DOCUMENT RESUME

ED 195 338 PS 011 848

AUTHOR Ruopp, Richard; And Others

TITLE Technical Appendices to the National Day Care Study: Measurement and Methods. Final Report of the National Day Care Study. Volume IV-B.

INSTITUTION Abt Associates, Inc., Cambridge, Mass.

SPONS AGENCY Administration for Children, Youth, and Families (DHEW), Washington, D.C.

PUB DATE Oct 80

CONTRACT 105-74-1100

NOTE 340p.; For related documents, see ED 131 928-930, ED 147 016, ED 152 114, ED 160 188, ED 168 706, ED 168 733, and PS 011 846-849.

EDRS PRICE MF01/PC14 Plus Postage.

DESCRIPTORS *Classroom Environment; Classroom Observation Techniques; *Day Care; Day Care Centers; Early Childhood Education; Economic Factors; Measurement Techniques; *Measures (Individuals); National Surveys; Policy Formation; Preschool Children; *Psychometrics; *Research Methodology

IDENTIFIERS *National Day Care Study

ABSTRACT

This volume, "Measurements and Measures" is the second of three Technical Appendices to the National Day Care Study (NDCS) and presents seven individual reports on a series of technical tasks undertaken to support the principal analyses of the effects of key center characteristics on children. Among the topics covered: analysis of alternative measures of classroom composition; psychometric analysis of the NDCS test battery; and analyses of several other more peripheral instruments used in the study. Also presented are results of a special survey of parents of subsidized children taken during Phase III of the NDCS, and analyses of the impact on children of other center characteristics, such as physical space and program orientation, and economic factors. (Author/MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195338

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Final Report of the National Day Care Study
VOLUME IV-B

October 1980

TECHNICAL APPENDICES
TO THE NATIONAL
DAY CARE STUDY

MEASUREMENT AND METHODS

Papers by:

William Bache III
Nancy N. Goodrich
Jean Layzer
Barbara Dillon
Daniel Calore

Prepared for:

Day Care Division
Administration for Children, Youth and Families
Office of Human Development Services
Department of Health, Education and Welfare

Allen N. Smith, Government Project Director

Prepared by:

Abt Associates Inc.
Cambridge, Massachusetts

Richard Ruopp,
National Day Care Study Director

Jeffrey Travers,
Associate Director

Nancy N. Goodrich,
Technical Coordinator

Contract No. 105-74-1100

PS011848

OVERVIEW OF NDCS FINAL REPORT VOLUMES

Results of the National Day Care Study and its major supporting study, The National Day Care Supply Study, is presented in a five-volume final report. Contents of these volumes are as follows:

Volume I

Children at the Center: Summary Findings and Policy Implications of the National Day Care Study presents in summary form the major findings and implications for federal day care policy of the National Day Care Study, a four-year study of the effects of regulatable center characteristics on the quality and cost of day care for preschoolers. Volume I serves both as a self-contained volume for the policy makers and as the foundation for the detailed presentation of results in Volumes II, III and IV. (Executive summaries of Supply Study findings and findings of an Infant/Toddler Study are included as appendices to Volume I.)

Volume II

Research Results of the National Day Care Study is a companion volume to *Children at the Center*. Volume II documents the analyses and results of the NDCS for the technical reader who seeks a more thorough understanding of the study from a research perspective. Volume II thus provides the quantitative support for the findings and policy conclusions reported in *Children at the Center*.

Volume III

Day Care Centers in the U.S.: A National Profile 1976–1977, the final report of the National Day Care Supply Study, is based on data gathered from a national random sample of over 3000 day care centers, stratified by state. Summary information is presented on characteristics of children and families served, center programs, staff, finances and regulatory compliance. Discussion of results is augmented by over 150 statistical tables.

Volume IV

Technical Appendices to the National Day Care Study is a compendium of technical papers supporting the most important conclusions of the study. These papers form the basis for the summaries in Volumes I and II. NDCS appendices are bound in three sections as follows.

Volume IV-A, *National Day Care Study Background Materials*, contains three papers, each of which establishes a distinctive context for the NDCS: a literature review focused on effects of group care and regulatable characteristics of the day care environment; case studies of the history and current practice of day care in the three NDCS sites (Atlanta, Detroit, Seattle); and a review of child development issues relevant to the NDCS from the perspective of black social scientists.

Volume IV-B, *National Day Care Study Measurement and Methods*, presents individual reports on a series of technical tasks supporting the principal analyses of the effects of key center characteristics on children. Among the topics covered are: analysis of alternative measures of classroom composition; psychometric analysis of the NDCS test battery; and analyses of several other more peripheral instruments used in the study. Also presented are results of a special survey of parents of subsidized children taken during Phase III, analyses of the impact on children of other center characteristics, such as physical space and program orientation, and econometric analyses.

Volume IV-C, *National Day Care Study Effects Analyses*, also a series of individual technical reports, begins with a presentation of the major effects analyses based on the two behavioral observation instruments, and then moves to a detailed treatment of the development and use of adjusted test score gains. The links among caregiver and child behavior, child test scores and other dependent measures are explored. Also detailed are results of the Atlanta Public School (APS) controlled substudy and APS replication substudy.

Volume V

National Day Care Study Documentation and Data gives a brief overview of NDCS data collection instruments and data files. Part A consists of the instruments themselves, including interview and data collection forms, observation systems and cognitive tests. Part B consists of data dictionaries; these describe every variable in the NDCS analytic data files. Part C provides codebooks for the data files. Parts B and C are available on computer tapes, which are readable independent of specific computer systems. Note that computer tapes are available only from Abt Associates.

Copies of the final report may be ordered from:

- EXECUTIVE SUMMARY (ONLY)
  Day Care Division
  Administration for Children, Youth and Families
  Office of Human Development Services
  Department of Health, Education and Welfare
  400 6th Street, S.W.
  Washington, D.C. 20024

- EXECUTIVE SUMMARY, Volumes I–IV
  ERIC Document Reproduction Service
  Computer Microfilm International
  P.O. Box 190
  Arlington, VA 22210

- EXECUTIVE SUMMARY, Volumes I–V
  Abt Associates Inc.
  55 Wheeler Street
  Cambridge, MA 02138

Earlier NDCS publications available from ERIC (hard copy or microfiche) are:

National Day Care Study First Annual Report, Volume I: An Overview of the Study [order number ED 131 928], *Volume II: Phase II Design* [order number ED 131 929], and *Volume III: Information Management and Data Collection Systems* [order number ED 131 930] (Cambridge, MA; Abt Associates, 1976).

National Day Care Study Second Annual Report [order number ED 147 016] (Cambridge, MA; Abt Associates, 1977).

National Day Care Study Preliminary Findings and their Implications [order number ED 152 114] (Cambridge, MA; Abt Associates, 1978).

PAPERS:

Comparing Alternative Measures
of Classroom Composition
(1)

A Psychometric Analysis of the
National Day Care Study Phase III
Child Test Battery
(51)

Investigation of Teacher Rating Scales
Considered for Use in
the National Day Care Study
(103)

An Analysis of the CDA Checklist Data
(119)

Interviews with Parents
(185)

The Classroom Environment Study
(217)

The Econometric Model
(261)

GLOSSARY

This glossary is intended as an aid to the reader. It is not an exhaustive dictionary of terminology relevant to the study or practice of day care, but rather a list of terms used throughout the volume which may be unfamiliar to the reader or which have special meanings for the purposes of the National Day Care Study.

An alphabetical list of terms enables the reader to find any item easily; numbers refer to the location of the term in the glossary itself, which is arranged by subject area to facilitate understanding of terms in relation to each other and in the context of this study. Subject areas are:

Classification of Day Care Services
Children and Staff
Classification of Day Care Centers
NDCS Independent Variables
NDCS Dependent Variables
Statistical Terminology

Alphabetical List of Terms

activity subgroup [42]
aide [17]
auspices [21, 25]
background variable [46]
caregiver [13]
caregiver/child ratio [44]
caregiver qualifications [45]
child outcome [51]
classroom composition [38]
classroom process [49]
core care [8]
correlation [59]
cost variables [54]
day care [1]
day care center [2]
dependent variable [47]
developmental outcomes [52]
effects [48]
family day care home [3]
FFP center [34]
full-time day care [6]
funding source [30,33]
generalizability of a measure [57]
generalizability of a sample [58]
group center [23]
group day care home [4]
independent center [22,26]
independent variable [36]
infant [12]
in-home day care [5]
lead caregiver [16]
lead teacher [15]
legal status [19]
multiple regression [61]

non-FFP center [35]
nonprofit center [24]
number of caregivers [39]
outcome [53]
parent-fee
part-time day care [7]
policy variable [37]
preschooler [10]
principal components analysis [62]
private center [28]
process [50]
profit center [20]
provider [18]
public center [29]
publicly funded center [32]
regression [60]
reliability [56]
sponsored center [27]
staff [14]
staff/child ratio [43]
staffing pattern [40]
supplemental services [9]
toddler [11]
validity [55]

Classification of Day Care Services

Day Care [1] is defined as care provided to a child by a person or persons outside the child's immediate family, either inside or outside the child's home.

- A day care center [2] is defined as a licensed facility in which care is provided to 13 or more children under the age of 13, generally for up to 12 hours each day, five or more days each week, on a year-round basis.

- The term family day care home [3] refers to a private family home, generally not licensed, in which children receive care, usually for up to 12 hours each day, five or more days each week, on a year-round basis. Most state licensing codes limit family day care homes to a maximum of six children.

- A group day care home [4] is defined as a private home serving 7 to 13 children, with one or two adults.

- In-home day care [5] is defined as care provided to a child in the child's own home by a nonrelative or by a relative who is not a member of the child's immediate family.

Day care of any of these types may be either full-time or part-time.

- Full-time day care [6] is defined as care for 30 or more hours per week.

- Part-time day care [7] is defined as care for less than 30 hours per week.

The services provided by a day care center may be classified into two blocks.

- Care care [8] refers to the common components of the daily experience of all children in day care centers. Core care includes provision of meals, snacks, space and educational/play materials, arrangements for minimum health care, and various caregiver services necessary to the nurturance of young children.

- Supplemental services [9] are those services to children and their families provided by a day care center in addition to core care. For children, such services include transportation, diagnostic testing and referrals. For parents, examples are social, welfare and employment services, and parent involvement in advisory and decisionmaking capacities. Supplemental services often address fundamental needs; the term "supplemental" merely reflects the fact that they are outside the scope of a minimal center day care program.

Children and Staff

The following terms are applied to children and adults in day care settings.

- Preschoolers [10] are defined as children three, four and five years of age (36-71 months). In some states most five-year-olds attend kindergarten and thus are considered school-aged children. In these cases, preschoolers are predominantly 36 through 59 months of age.

- _Toddlers_ [11] are defined as children aged 18 through 35 months of age.

- _Infants_ [12] are defined as children from birth through 17 months of age.

- _A caregiver_ [13] is a person who provides direct care to children in a day care center classroom, a family day care home, or in a child's own home. Unless otherwise specified, the terms _caregiver_ and _staff_ [14] are interchangeable in NDCS documents.

- _A lead teacher_ [15] (or _lead caregiver_ [16]) is the principally responsible caregiver in a day care classroom. The term "teacher" is not intended to connote a school-like atmosphere in the day care center. The term _caregiver_ has been used to refer to persons working with children in day care settings, and the term _lead teacher_ is sometimes used to distinguish the principally responsible caregiver in a day care classroom from her aides.

- An _aide_ [17] is a caregiver who assists a lead teacher in a day care classroom.

- A day care _provider_ [18] is a person who is directly or indirectly involved in the provision of day care services; including caregivers, center directors and owners.

Classification of Day Care Centers

Day care centers are classified according to _legal status_ [19] as profit or nonprofit.

- _Profit centers_ [20] are further classified according to auspices [21] as independent centers or group centers.

  --_Independent centers_ [22] are not part of a chain of day care centers.

  --_Group centers_ [23] belong to a chain (group) of day care centers.

- Nonprofit centers [24] are classified according to auspices [25] as independent centers or sponsored centers.

  --Independent centers [26] are not sponsored by any group or agency.

  --Sponsored centers [27] are classified as either private or public, according to the nature of the sponsoring agency.

    --Private centers [28] are sponsored by a private agency, such as a church. (Note that all profitmaking centers, as well as independent nonprofit centers, are necessarily private.)

    --Public centers [29] are sponsored by some government agency, such as a city school system or a county welfare department.

In addition to classification by legal status and auspices, day care centers may be classified by a crosscutting typology according to funding source. [30]

- Parent-fee centers [31] derive more than half of their income from parent fees.

- Publicly funded centers [32] derive their funding principally from government subsidies and gifts and contributions.

Alternatively, centers may be classified by funding source [33] according to federal financial participation (FFP). This typology was used in Supply Study analyses, and the reader may find these terms used when Supply Study data are referred to.

- An FFP center [34] is defined as any center which serves one or more federally subsidized child(ren).

- A non-FFP center [35] is defined as a center which serves no federally subsidized children.

NDCS Independent Variables

NDCS independent variables [36] are those variables whose costs and effects were to be measured. There are two types of independent variables: policy variables and background variables.

- Policy variables [37] are those characteristics of day care centers which may influence the quality and cost of center day care and which are or can be affected by federal policy. The NDCS was concerned with two major classes of policy variables: classroom composition and caregiver qualifications:

  --Classroom composition [38] describes configurations of caregivers and children in day care classrooms. Classroom composition is defined by three variables. (Note that any two of these variables mathematically define the third.)

    --Number of caregivers [39] is defined as the total number of caregivers assigned to each classroom. (The term staffing pattern [40] may refer not only to the number of caregivers assigned to a classroom, but also to the mix of teachers and aides or to the mix of qualifications of the caregivers in a classroom.)

    --Group size [41] is defined as the total number of children assigned to a caregiver or team of caregivers. In most cases, groups occupied individual classrooms or well-defined physical spaces within larger rooms. In a few "open classroom" centers, children were free to move from group to group. In such cases, clusters of children participating in common activities under the supervision of the same caregiver or team of caregivers were considered to be "groups." (The term activity subgroup [42], by contrast, refers to the actual number of children interacting with a particular caregiver. A group of 20 children, for instance, might be divided into three activity subgroups, one with the lead teacher, and two with aides.)

--Staff/child ratio [43] is defined as number of caregivers divided by group size. Higher, or more stringent, staff/child ratios are those with a smaller number of children per adult. For instance, a ratio of 1:5 is higher, or more stringent, that a ratio of 1:10 (which is lower, or less stringent). Note that the terms staff/child ratio and caregiver/child ratio [44] are interchangeable in NDCS discussions.

--Caregiver qualifications [45] variables were developed to describe caregivers' years of formal education, amount of training and/or education related to child development, and amount of work experience as a caregiver.

- Background variables [46] are characteristics of day care centers which can be influenced by government regulation only indirectly, if at all. Examples are age, sex and race of children, or socio-economic characteristics of families and of the community served by a center.

NDCS Dependent Variables

NDCS dependent variables [47] are those features of day care costs and quality measured as indicators of the effects of such center characteristics as group size, staff/child ratio and caregiver qualifications (the study's independent variables).

- In NDCS discussions, the term effects [48] is often used to distinguish dependent variables pertaining to quality in day care from dependent variables pertaining to day care costs. There are two major classes of effects variables.

  --The term classroom process [49] (or process [50]) refers to the behavior of children and caregivers in the classroom; that is, the dynamics of their interaction. Process was recorded using two observation instruments, one concentrating on children's behaviors (the Child-Focus Instrument) and one concentrating on caregivers' behaviors (the Adult-Focus Instrument).

- --The term child outcomes [51] (or developmental outcomes [52], or outcomes [53]) refers to children's gains in school-readiness skills; although a number of tests and ratings of social and cognitive development were field-tested, ultimately only two, both standardized cognitive tests, proved reliable enough to be used as outcome measures: the Preschool Inventory (PSI) and the Peabody Picture Vocabulary Test (PPVT).

- Cost variables [54] correspond in the main to commonly used terminology in accounting and economics. Where terms or variables peculiar to the NDCS are introduced, they are explained in the text.

Statistical Terminology

- The validity [55] of a measure is the degree to which it measures what it purports to measure. Various features of a measure may be indicative of its validity; such as: (1) a direct conceptual relationship between the measure and the construct of interest (e.g., between an observer's count of the number of children present in a class and the variable group size); or (2) agreement with other measures of the same construct (e.g., agreement between observation-based measurements of group size and schedule-based measurements of group size).

- The reliability [56] of a measure is the degree to which it gives consistent results when applied in a variety of situations; that is, the degree to which it is free of measurement error. Reliability coefficients vary from 0.00 to 1.00. A coefficient of 0.00 indicates a completely unreliable measure; a coefficient of 1.00 indicates a measure that gives perfectly consistent results across all situations. Thus, a reliability coefficient of .95 indicates that 95 percent of the measured variation among the objects of measurement (e.g., among children) is attributable to genuine differences among the objects of measurement, and that only 5 percent of the variation measured is attributable to random effects of errors of measurement.

- The generalizability of a measure [57] is a sophisticated extension of the concept of reliability in psychological measurement theory. It incorporates the notion that the numerous sources of variation in measurement groups as "measurement error" according to standard reliability theory may or may not be defined as "error," depending on one's purpose in using a given measure. [The concept of generalizability is a very complex one which cannot be clearly presented in the limited space available here. For a definitive treatment of the subject, the reader is referred to L. Cronbach, G. Gleser, H. Nanda, and N. Rajaratnam, The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles (New York: John Wiley & Sons, Inc., 1972).]

- The generalizability of a sample [58] is the degree to which the sample accurately represents a universe to which findings based on the sample are to be extended.

- The correlation [59] (degree of association) between two variables is represented by a correlation coefficient expressed as a decimal fraction. Correlation coefficients range from +1.00 (representing a perfect positive correlation) through zero (representing the absence of any correlation) to -1.00 (representing a perfect negative correlation). For example, a positive correlation between children's scores on Tests A and B would mean that children with high (or low) scores on Tests A also tend to have high (or low) scores on Test B. If the two tests' scores were negatively correlated, then high scores on Test A would tend to be associated with low scores on Test B, and vice versa.

- Regression [60] analysis is a technique for extracting from data an idealized representation, in the form of a straight line, of the relationship between two variables. That is, regression defines the particular straight line which is the "best" linear approximation of the less clearcut pattern exhibited in the data. Similarly, multiple regression [61] analysis extracts an idealized representation of the relationships between a given dependent variable and two or more independent variables.

- Regression [60] analysis is a technique for extracting from data an idealized representation, in the form of a straight line, of the relationship between two variables. That is, regression defines the particular straight line which is the "best" linear approximation of the less clearcut pattern exhibited in the data. Similarly, multiple regression [61] analysis extracts an idealized representation of the relationships between a given dependent variable and two or more independent variables.

- Principal components analysis [62] produced alternative weighted combinations of variables ("principal components"), thus allowing the researcher to select a small number of components which convey most of the important information in a data set--that is, which together account for a large proportion of the variance in the data. For example, a large number of variables related to socioeconomic status might be reduced to a few components--clusters of variables which are highly correlated with one another and only weakly related to variables in other components.

FOREWORD

Providing sound research which supports social policy directions affecting the lives of children and families is unquestionably a major goal of the Administration for Children, Youth and Families. By producing a clear signal in an often times cloudy environment, we are able to fulfill this important responsibility that has been entrusted to us.

The National Day Care Study (NDCS) is an outstanding example of our meeting this responsibility. This study has been widely recognized in both public and private sectors as one of the most important social policy research investigations ever by the Department. Its information has been widely used by many people and organizations, and it already has had a major impact on the drafting of the new HHS Day Care Regulations.

The NDCS searched for day care center characteristics which can both protect children from harm as well as foster their social, emotional and cognitive development. It discovered that these outcomes are clearly attainable when groups of children are small and when caregivers receive training in child-related areas. It also found that relaxing the staff/child ratio would not adversely affect children but could lower costs substantially and thus enable more children to receive care. That these findings held up across diverse sites and with different groups of children, provided support that all children can benefit from a single set of standards.

In all, I feel that the NDCS has more than justified the tremendous energy and time that has gone into it. Through this kind of commitment to excellence in its research programs, the Administration for Children, Youth and Families can be an instrumental force in enhancing the well-being of all children and families.

I am pleased to present the final volumes of the study--Volumes II and IV-A, B and C. Volume II is the research companion to Volume I--"Children at the Center." It provides quantitative support to the study's major findings. Volume IV is a compendium of technical papers which address study-related background issues, NDCS measures and methods and detailed results of individual outcome areas.

Jack Calhoun
Commissioner, Administration
for Children, Youth and Families

October, 1980

PREFACE

The federal government has become a major purchaser of child care, chiefly for the children of the working poor. With the growth of federal expenditures has come increased public concern about the quality and cost of care purchased with federal dollars. The National Day Care Study (NDCS) addressed this dual concern. Commissioned in 1974 by the Office of Child Development,* the study was conducted by two private research organizations--Abt Associates Inc. and SRI International. The study concluded that, by setting appropriate purchasing standards, the government could buy better care at lower cost than it currently buys, thus allowing it to serve more children within existing budgets.

Results of the study were summarized in a report published in March 1979.[1] The results were heavily cited in supporting arguments for proposed federal regulations, which were published in the Federal Register in early 1980.[2]

The present volume is one of a series supplementing the summary report.[3] It is intended to provide professionals in developmental psychology and related fields with a description of the methods and findings underlying the study's conclusions about links between regulatable characteristics of day care centers and the experiences and development of preschool children in center care.

Policy Context of the NDCS

Public concern with the quality of federally subsidized child care is embodied in the Federal Interagency just mentioned), there was little evidence of major heterogeneity that might suggest that the effects of group size are site-specific. Moreover, there was no clear numerical point of demarcation between small, "good" groups and large, "bad" ones. Most of the study's centers maintained groups of three- and four-year olds that varied in size from 12 to 24; typically, desirable behaviors decreased in frequency by roughly 20 percent, and undesirable behaviors increased by 20 percent, as group size increased within this range.

---

*The Office of Child Development is now the Administration for Children, Youth and Families (ACYF).

Third, staff/child ratio was also related to some aspects of interaction in the classroom, but the correlates of this critical policy variable, the focus of much of the controversy surrounding day care regulations, were less widespread than those of group size. Ratio was most clearly related to caregiver behavior: lead caregivers in high-ratio classes (those with few children per adult) showed essentially the same pattern of behavior reported above for caregivers in small groups. (However, the confounding of ratio and group size for the lead caregiver sample made it unclear whether the behavior pattern should be attributed to ratio, group size or both.) In addition, lead caregivers in high-ratio classes spent less time in overt management of children than those in low-ratio classes. They also spent more time interacting with other adults and in other activities not directly involving children. Thus some of the "contact time" potentially available to children by virtue of high adult/child ratios was spent in other ways. High ratios were not associated with high frequencies of one-to-one interaction between adults and children; in fact, ratio showed few systematic relationships to the behavior of children at all. Nor was ratio related to children's test score gains, except in a few isolated instances.

Title XX FIDCR. That report, issued in 1978, concluded that federal regulation was an appropriate means of maintaining quality in subsidized care but that the existing FIDCR were in need of revision.[5]

The Office of Child Development (now ACYF) had initiated the NDCS before the controversy over the Title XX FIDCR erupted. The NDCS and the Appropriateness Report were entirely independent efforts. Nevertheless the authors of the Appropriateness Report made heavy use of early results from the study, incorporating a preliminary report of NDCS findings [6] as an appendix to their own report. Subsequently, NDCS staff and the government project director were consulted during the drafting of revised regulations, which began within ACYF and was completed by the Office of HEW's General Counsel. The influence of the study is clearly visible in the proposed new standards regarding caregiver qualifications and group composition (group size and staff/child ratio). While the proposed standards deviate from the specific numerical recommendations regarding ratio and group size that appeared in the NDCS 1979 summary report, basic principles are retained--notably joint regulation of ratio and group size, with increased emphasis on the latter--as are many detailed suggestions regarding methods of monitoring and enforcement.

NDCS Approach and Findings: An Overview

The 1968 FIDCR were based on the advice of practitioners and experts in fields related to child care, as well as the best research evidence available at the time. However, in 1968 there existed only limited empirical evidence to support the basic but tacit assumptions that link various provisions of the regulations to quality of care--for example, the assumption that maintaining high staff/child ratios (few children per caregiver) will increase the quantity and quality of adult-child interaction. Nor were there data to support the assumption that regulatory control over such center characteristics as staff/child ratio, group size and staff qualifications would produce similar outcomes for children across the regions, states, sponsoring agencies and socioeconomic groups affected by federal legislation. Similarly, though a good deal was known about the different components of cost in day care, no specific evidence existed to link costs to regulated center characteristics or to quality. The NDCS attempted to fill these gaps in knowledge by identifying costs and effects associated with variations in center characteristics that were regulated or could potentially be regulated by the federal government.

The study's sponsors and designers recognized that national policymakers have many different views of the goals of day care. For example, federally subsidized day care can be seen primarily as an institution designed to free parents to work or to employ welfare recipients. However, ACYF has long been committed to the view that day care can and should foster the development of children. Hence the study focused on the quality of care from the point of view of the child--i.e., on the nature of the child's experience in day care and on the developmental effects of that experience, as measured by naturalistic observations and standardized tests. While many potentially regulatable center charcteristics were examined, primary attention focused on those characteristics which seemed most central to existing regulations and most likely to affect the daily experience of the child, namely staff/child ratio, group size and staff qualifications.

Perhaps the most general and important finding of the study was that variations in regulatable center characteristics do make a difference in the well-being of children. In contrast to many earlier studies of the effects of variations in curriculum or resource outlay in education, the NDCS showed clearly that it matters how day care classes are arranged and who staffs them. To be sure, much of what goes on in day care is not influenced by regulatable center characteristics. There is a great deal of variability in the quality of human interaction in day care settings even when the composition of the classroom and the qualifications of caregivers are fixed. Nevertheless regulatable characteristics show relationships to meaures of children's experience and of developmental change that are significant both statistically and substantively.

More specifically, for preschool children (ages 3-5), the smaller the group in which children are placed, the more they tend to engage in creative, verbal/intellectual and cooperative activity. Also, children in small groups make more rapid gains on certain standardized tests than do their peers in larger groups. When groups are larger, individual children tend to "get lost," i.e., to wander aimlessly and to be uninvolved in the ongoing activity of the group. These findings hold even when staff/child ratios are relatively high (i.e., when there are few children per caregiver).* Adding adults (usually teachers' aides) to a large group of children improves the adult/child ratio but does not necessarily result in increased engagement on the part of the child, nor improved test score gains. Significantly, children do not appear to experience more one-to-one interaction with adults when ratios are high than when they are low.

---

*In day care classrooms, unlike many public school classrooms, it is not usual to find a single adult in charge. Configurations of two or three caregivers, usually a teacher plus aides, are more common. Both the number of children and the number of adults varies significantly from classroom to classroom. It is for this reason that staff/child ratio and group size can vary more or less independently and must be examined separately. It cannot simply be assumed that large classes will have low ratios nor that small classes will have high ratios.

The behavior of caregivers toward children is also related to group or class size, but it is related to the staff/child ratio as well. In small classes and/or classes with high ratios (few children per caregiver), staff tend to devote their attention to small clusters of 2-7 children, rather than to large clusters of 13 or more. Staff in such classes also spend less time observing children passively than do caregivers in large classes and/or classes with low ratios. In addition, the staff/child ratio shows some relationships to caregiver behavior that are not found for group size. High ratios appear to make management of children easier. Also, in high-ratio classes adults spend more time with other adults and in activities not involving children, such as performance of routine chores. This outcome may suggest that high ratios benefit caregivers by providing contact with other adults and time to do necessary tasks, but it also suggests one reason why high ratios do not appear to affect the amount of one-to-one interaction between caregivers and children: in high-ratio classes some of the time potentially available for children is diverted to activities in which children are not directly involved.

On balance, NDCS findings suggest that the importance of group size as a regulatory device for influencing quality in child care may have been underestimated and the importance of staff/child ratio somewhat overestimated. This conclusion, of course, is not an argument for abandoning regulation of staff/child ratio. Not only did ratio show some positive effects, but the range of ratios examined in the NDCS was relatively narrow and relatively high. (Most centers in the study maintained classes with five to nine children per caregiver.) This range was chosen to illustrate effects of variations in ratio between levels required by the FIDCR and levels permitted by most states. Consequently, generalization of the findings to levels outside the range established by current regulatory variations is unwarranted. Moreover, a subsidiary study of center care for children under three suggested that ratio was as important as group size in influencing quality of care for infants and toddlers. Thus, while the findings suggest that controlling ratio alone is not an effective regulatory strategy, they also suggest that ratio should be included with group size in regulations governing classroom composition.

In addition to the above findings on group composition, the NDCS showed that qualifications of caregivers also affect quality of care. While years of formal education, degrees attained and years of experience per se made no discernible difference in quality of care, those caregivers who had education or training specifically related to young children (e.g., in early childhood education, day care, special education or child psychology) provided more social and intellectual stimulation to children in their care than did other caregivers, and the children scored higher on standardized tests.

To arrive at policy recommendations, these findings were integrated with results from other components of the study which were concerned with the costs associated with the various regulatable center characteristics and with prevailing practices in staffing and group composition among centers nationally. The costs of maintaining small groups and of employing staff trained or educated in child-related fields were found to be small, whereas the costs associated with maintaining high staff/child ratios were significant. Consequently it was recommended that, for preschoolers, the group size standards of the existing FIDCR be maintained or made more stringent, while the ratio requirements be relaxed slightly. The expected result would be an improvement in the quality of care for preschoolers together with a reduction in costs relative to those that would prevail if the Title XX FIDCR were enforced. Implementation of the NDCS recommendations would not require major disruption of current practice, since a high proportion of centers nationally already maintain both relatively small groups and staff/child ratios that are only a little less stringent than those mandated by the FIDCR,* despite claims of some providers and state Title XX administrators that the FIDCR ratios are unrealistically strict.[7] For infants and toddlers, institution of a group size standard and maintenance of the current ratio standard were recommended. It was also recommended that training or education in a child-related field be required of all individuals providing direct care to children, and that states be required to make such training available.

Organization of Technical Appendices

Technical Appendices to the National Day Care Study are divided into three volumes. Volume IV-A, *Background Materials*, contains three papers that help to set a context for overall study results: "Research Issues in Day Care, A Focused Review of the Literature," "Case Studies of the National Day Care Study Sites: Atlanta, Detroit and Seattle," and "The National Day Care Study from the Prospective of Black Social Scientists: Reflections on Key Research Issues." Volume IV-B *Measurement and Methods* provides seven papers that describe technical tasks undertaken to support the effects analyses reported in Volume IV-C. Included are papers about "Comparing Alternative Measures of Classroom Composition," "A Psychometric Analysis of the National Day Care Study Phase III Child Test Battery,"

---

*Staff/child ratios nationwide, averaging over all classes and ages of children, are 1:6.8, compared to 1:6.3 required by the FIDCR, and 1:12.5 permitted by state licensing requirements.[8]

"Investigation of Teacher Rating Scales Considered for Use in the National Day Care Study," "An Analysis of the CDA Checklist Data," "Interviews with Parents," "The Classroom Environment Study," and "The Econometric Model."

Volume IV-C, Effects Analyses, presents the results of analyses that investigated relationships between policy variables, classroom processes and child outcomes. Six papers are included: "The Adult-Focus Observation Effects Analysis," "The Child-Focus Observation Effects Analysis," "Analysis of Test Score Growth in the National Day Care Study," "Classroom Process-Child Outcome Analyses," "The Atlanta Public Schools Day Care Experiment," and "The Effects of Day Care in Eight Atlanta Public Schools Day Care Centers." All of the papers in the Technical Appendices were prepared by study analysts and were the basis for findings presented in Volumes I and II.

PREFACE References

1. Ruopp, R., Travers, J., Glantz, F., and Coelen, C. Children at the Center. Final Report of the National Day Care Study: Summary Findings and their Implications. Cambridge, Mass.: Abt Books, 1979.

2. Federal Register, March 19, 1980.

3. Other supporting volumes include Coelen, C., Glantz, F., and Calore, D. Day Care Centers in the U.S.: A National Profile 1976-1977. Cambridge, MA: Abt Books, 1978; and three volumes of Technical Appendices to the National Day Care Study. Cambridge, Mass.: Abt Associates Inc., 1980.

4. Ruopp, et al., op. cit., Appendix A. p. 231.

5. Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare. The Appropriateness of the Federal Interagency Day Care Requirements: Report of Findings and Recommendations. Washington, D.C.: U.S. Government Printing Office, 1978.

6. Travers, J., and Ruopp, R. National Day Care Study Preliminary Findings and Their Implications Cambridge, Mass.: Abt Associates Inc., 1978.

7. See Ruopp, et al., op. cit., Chapter 8, 155, and Appendix A, 230-240.

8. Ruopp, et al., op. cit., Appendix A., 236.

ACKNOWLEDGEMENTS

Since the National Day Care Study (NDCS) began in 1974, a great number of people have participated in the effort. These include project staff at Abt Associates Inc., site staff in Atlanta, Detroit and Seattle, and a panel of consultants from across the country, all of whom were ably directed by Dr. Richard Ruopp, Project Director. Staff and consultants at the Administration for Children, Youth and Families, and in particular Mr. Allen Smith, the Government Project Officer, also provided valuable direction for the study. Individual staff and roles are acknowledged in greater detail in Volume I, Children at the Center.

The final task for the NDCS has been the preparation of these volumes of Technical Appendices. The authors of the papers contained herein wrote, rewrote and revised their individual papers to prepare them for publication. In many instances Dr. Lorelei Brush and Dr. Jeffrey Travers gave special technical direction and Sally Weiss provided editorial assistance. Nonetheless, each paper represents an individual effort by each author. No attempt has been made to ensure consistency of style or format or to link the findings of the various papers; this was the purpose of Volume I, Children at the Center and Volume II, Research Results.

Producing all of these papers has required a considerable effort by Karen Hudson, secretary for this final phase of the study, and Christine Bornas, secretary during the earlier phases. They have managed to prepare drafts, organize changes, make corrections and produce the final papers, always within the time schedules provided. To them, and to all of the authors and support staff, I give my warmest thanks.

Nancy Goodrich
Technical Coordinator
Abt Associates Inc.
Cambridge, Mass.
October, 1980

Comparing Alternative Measures of Classroom Composition

William L. Bache III

TABLE OF CONTENTS

Page

SUMMARY 3

CHAPTER ONE: INTRODUCTION TO THE PROBLEM 6

CHAPTER TWO: MEASURING ROSTERED, SCHEDULED, AND OBSERVED CLASSROOM COMPOSITION IN THE NDCS 13

- Center Audits of Staffing and Enrollment 13
- Center Inspections of Classroom Composition 17
- Concerning the Use of Staff/Child, Child/Staff, and Long-Transformed Ratios 23

CHAPTER THREE: COMPARISONS AMONG ROSTERED, SCHEDULED AND OBSERVED CLASSROOM COMPOSITION 27

- The Influence of Merging on Classroom Composition Measures 28
- The Effect of Absenteeism on Scheduled Composition 33
- The Effect of Volunteerism on Caregiver Measurement 36

CHAPTER FOUR: THE FEASIBILITY OF MONITORING REGULATORY COMPLIANCE WITH OBSERVATIONS 41

- Conclusions 47

SUMMARY

This paper compares three different techniques for measuring classroom composition (defined as group size, number of caregivers, and the resulting caregiver/child ratio) in day care classrooms. Two of these are audit-oriented methods based on records reflecting a center's plan for caregiving: a roster system using head counts, irrespective of full- or part-time status, and a schedule system which computes full-time equivalents (FTE's) from weekly hours of work or attendance. The third method, using actual observations of classrooms to obtain counts of children and caregivers, is inspection oriented and reflects the care actually delivered by a center.

Whether or not a class tended to merge occasionally with one or more other classes was found to have a profound effect on the relationship between audit and inspection measures of classroom composition. Among classes which rarely merge, observed group size and caregiver count are nearly identical to the corresponding schedule measures if the latter are adjusted for absenteeism. If no adjustment is made for absententeeism, scheduled FTE's exceed observed counts. Among classes which tend to merge, however, observed group size is approximately eight children larger than absentee-adjusted scheduled group size and the observed caregiver count is greater than the scheduled by nearly one person. Observed and scheduled ratio, however, are not significantly different in merging classes and are more strongly related to each other than they are in nonmerging classes.

Merging is a phenomenon which may be planned in advance but often is not. In the five Phase III rounds of inspections, the proportion of NDCS target classes which

were observed to merge somewhat regularly in the morning varied from 16 to 28 percent. (Many centers merged classes during lunch or nap, but these times of the day were not included in any of the observations analyzed for this paper.) Moreover, not only the number, but the identity of centers which merged classes varied during the course of the year. Although 24 centers were observed to merge classes in at least one of the inspection rounds, only ten did so in three or more rounds and only six did so in all five rounds. Twelve centers, on the other hand, merged classes only once in five rounds of observations spread across the year. (The number of centers which merged classes in any given round varied from a low of 10 in October and January to a high of 16 in April.)

The tendency to merge classes, therefore, is a stable center characteristic only in some centers. In the majority of centers which merge classes at all, it appears to be a relatively unpredictable event. Even in centers where merging does occur regularly throughout the year, it is not necessarily planned in the sense that it can be indicated in center schedules prepared in advance.

Independent of the issue of merging, audit measures of classroom composition were also noted to be inappropriate for centers with an unstructured, open-classroom environment. Whereas observations can measure the composition of naturally-occurring but unscheduled groups in such situations, rosters and schedules must revert to describing the enrollment and staffing of the center as a whole.

In general, analyses comparing center-prepared audit records with agency-conducted inspections clearly indicate that schedules alone cannot validly reflect configurations of children and caregivers actually occurring in day care classrooms. Only observations appear capable of dealing with unplanned, extemporaneous events that characterize the day care environment. Finally, the logistics of conducting such observations may not be prohibitive for overseeing agencies. Variance components analyses indicate that a single morning visit to a center is sufficient to obtain a reasonably accurate estimate of the average classroom composition in effect during the week or perhaps even the month in which the visit is made.

Whether the accuracy obtained from one visit is sufficient for deciding whether a center is in compliance, however, depends upon the decision rules specified in the regulatory guidelines. The guidelines should state the span of time to which compliance measurements must generalize and with what degree of accuracy. It may, for example, be felt that compliance decisions should be based on a center's average performance over the course of a year rather than its situation in any given week.

NDCS data are insufficient to determine the accuracy obtainable from a single visit or a single schedule in estimating the average composition in a center over the course of a year. If the year is determined as the basis on which compliance decisions must be made, however, it is likely that information would have to be collected at several time points to achieve reasonable accuracy. Some mixture of audits and inspections may well prove desirable, given the advantages of each. A monitoring system, for example, in which occasional inspections were used to verify periodically-submitted schedule data or self-collected center observations would seem to be feasible.

CHAPTER ONE: INTRODUCTION TO THE PROBLEM

In the context of policy regulation, classroom composition may be defined as a set of three variables: group size, number of caregivers, and the resulting caregiver/child ratio. Only two of these need be regulated, however, since the third is determined once levels for the other two are established. Historically, regulations of classroom composition have focused on ratio and group size, the actual levels required being contingent on the age of the children being cared for. The 1968 Federal Interagency Day Care Requirements (FIDCR), for example, require a ratio of one caregiver to five childrenin groups not larger than 15 three- to four-year olds, and one caregiver to seven children in groups not larger than 20 four- to six-year-olds. Despite the explicitness of such regulations, however, classroom composition is not clearly defined by the FIDCR (nor by most state regulations) because the method of *measuring* ratio and group size is not clearly defined.

At the heart of this issue is the contrast which can be drawn between *audit-oriented* and *inspection-oriented* measures of classroom composition. Audit-oriented measures rely on records maintained by a center regarding child enrollment and staff assignments. The arrangement of children (and staff) in separate classes or groups and the age distribution of such groups may also be recorded. Audit-based measures can be further described as either *roster-derived* or *schedule-derived*.*

Rostered classroom composition is based on counts of the number of children and the number of staff assigned

---

*"Schedule" refers here to the *weekly* hours of attendance or work, not to a daily schedule of arrival and departure or even to a daily total.

to the classroom. The resultant variables are sometimes referred to as "head-count" measures. Group size and staff/child ratio, for example, may be determined from a simple center roster of all children and staff, provided the roster indicates the class assignment of each individual. Strictly speaking, no consideration is given to the number of hours each child and each staff person spends in the class, although a crude adjustment is sometimes made by dividing the number of part-time individuals by two before adding this to the number of full-time persons.

Schedule-derived classroom composition is based on the scheduled hours of contact between the children and the staff in a class, the variables being constructed from the children's attendance schedules and the staff's work schedules. Scheduled ratio, for example, is computed simply by dividing the total number of hours all caregivers are scheduled to spend in the classroom by the total number of scheduled hours of all children in the class. Group size and the number of caregivers, however, must be computed as a number of full-time-equivalent (FTE) persons by dividing the total number of hours by some constant (typically 40) which represents the weekly schedule for a full-time person. Scheduled composition variables are sometimes referred to as "contact-hour" measures, although it is important to remember that scheduled, rather than actual, contact hours are involved.

In essence, rostered and scheduled composition variables reflect the operational plan of care provided by the center. Unless detailed attendance records are also maintained, the classroom composition actually implemented cannot be determined precisely. Furthermore, the movements of children and staff in a day care center are not usually as controlled as they are, for example, in most schools. Such flexibility of movement, as well as the merging or splitting of groups for various activities, produces fluctuation in the actual classroom composition which even accurate attendance records cannot reconstruct.

Inspection-oriented measures of classroom composition, as the term implies, are based on actual visits to the centers being monitored. Observations of each classroom are conducted, during which counts of the children and staff present are obtained. Consequently, such variables may also be referred to as observation-derived measures. Inasmuch as observed classroom composition is based on a "snapshot" of the classroom as it appears at a given moment, adjustment for absenteeism is an intrinsic part of the measurement method. Observation measures also automatically adjust for variations in the work and attendance schedules of those in the class, provided that the observations conducted are representative of the class's schedule and activities. Although the number of observations required to achieve this representativeness is an empirical issue directly related to the feasibility of using observations to monitor compliance, the method itself possesses prima facie validity with respect to the classroom composition actually achieved by the center. In fact, the tradeoff between feasibility of administration and accuracy/validity is the central issue underlying the relative advantages and disadvantages of the audit and inspection methods.

On the one hand, audit-oriented methods appear more feasible because they do not necessarily require center visits by regulatory or licensing personnel. Child and staff rosters or schedules can be sent to the monitoring agency directly, simplifying the logistical aspects considerably. Data collection and processing can be centralized, thus potentially easing the burden, experienced by many agencies, of monitoring a large number of centers with too few personnel. As mentioned earlier, however, such data measure planned, rather than implemented, classroom composition. The difference between the two varies among centers, but can be considerable. If some classes merge for various activities or at various times of the day, the average observed group size can be much larger than the scheduled group size, which is based on the separate groups. Moreover, audited group size may be meaningless for centers with an "open classroom" or unstructured approach to child assignments. When children are permitted free access to any group or class and freedom of choice between ongoing activities, the only rostered or scheduled group size which is computable is that based on the center enrollment as a whole. When there is no _a priori_ disposition of children into groups or classes which is then followed in practice, there can be no rostered or scheduled group size other than that for the center as a single group. A final disadvantage of audit-oriented measures is that they are more easily distorted by centers desiring to do so. _Periodic visits to verify center-constructed records sent to the monitoring agency are the only way to authenticate the data received_. Yet such procedures erode the feasibility advantage typically advanced on behalf of the audit approach.

The principal advantage of the inspection method of measuring classroom composition is its _prima facie_ validity. It would seem self-evident that a count of the children and staff present in a group or room at a given moment is virtually a definition of the classroom composition at that moment. In particular, it may be noted that open-classroom centers do not pose the problem for measuring observed group size that they do for measuring rostered or scheduled group size. In such centers, children invariably form natural clusters, whether on their own or under staff guidance. The observation method can simply focus on such groups, counting the children and staff associated with each. Despite the fact that the number and composition of clusters in such centers is dynamic, the very nature of the observation method as a "snapshot" of a given moment allows their valid measurement in an unstructured setting.

The disadvantages of the inspection approach arise from the logistical problems of visiting every center to be monitored. One aspect of this problem, of course, is that inspections are qualitatively different from "passive" audits (in which the data are sent to the auditing agency), and require more planning and coordination. A more significant aspect, however, is the relationship between the number of observations required for accurate measurement (given fluctuations which occur during the day or across days) and the burden placed on available inspection personnel.

This does not imply that the required number is prohibitive or the burden exhaustive compared to audit methods, but merely that it is the most important logistical consideration. In fact, data will be presented later which indicate that a single center visit during the morning may be sufficient to obtain a reasonable estimate of the center's average composition for the month in which the visit is made. Moreover, monitoring procedures could be developed to lessen the burden further. Centers, for example, could collect their own observation counts once each month or two and send them to the agency. The agency might then need to conduct only one or two verification inspections yearly, with followup inspections of centers whose self-reported observations differed markedly from the verifications. Given the relatively high rate of FIDCR compliance reported by Coelen, Glantz and Calore (1978) during a year when the FIDCR were not even in force, such a model appears to be feasible.* Furthermore, centers might well prefer this approach, viewing it as less intrusive, less disruptive and less costly than audit procedures, which would require more personnel time to maintain the records being audited.

---

*According to this study, 76 percent of all non-profit centers receiving federal funds for one or more children complied with the FIDCR staff/child ratio requirements during 1976-1977, when the FIDCR had been suspended; 56 percent complied with the FIDCR group size requirement.

The comparative cost to monitoring agencies of audit versus inspection methods is not clear. The results would depend on specific details of the monitoring procedures--particularly the number of inspection visits. Potentially higher manpower needs for inspections, however, might be offset by higher manpower needs under audits to manage the larger data base and the more complicated data collection task.

Finally, certain issues regarding the measurement of classroom composition are independent of the methods discussed above, and must be resolved whatever monitoring procedure is used. Most notably, decisions must be made regarding which persons to include as caregivers and whether non-classroom time is to be included in computing the composition variables. Non-classroom volunteers and even aides could be excluded, as could the time spent by classroom staff outside the classroom (e.g., attending meetings, preparing meals, or filling out records). Inspection methods do not avoid the latter issue because decisions must be made concerning the inclusion of staff who happened to be momentarily absent at the time the observation count was made.

The remainder of this paper is devoted to statistical comparisons among rostered, scheduled, and observed ratio and group size. Section 2 describes the methods used in the NDCS to collect classroom composition data and to construct the variables. Section 3 presents the results of analyses comparing rostered, scheduled, and observed variables with each other. Finally, Section 4 presents

38

the results of variance components analyses which address the measurement accuracy obtainable with a single morning of center observation--and thus the feasibility of using observations to monitor compliance.

CHAPTER TWO: MEASURING ROSTERED, SCHEDULED, AND OBSERVED CLASSROOM COMPOSITION IN THE NDCS

The principal source of classroom composition data in the NDCS was a series of five audits and inspections conducted about every six weeks: in October, December, January, March and April, 1976-1977. In each center, both audit and inspection data were collected by an NDCS-paid secretary who worked half-time for the center and half-time for the study.

Center Audits of Staffing and Enrollment

Each of the five audits consisted of separate rosters of the children enrolled in each center and of the staff and volunteers working in the center. In addition to the rosters of names as such, other information was collected for each individual, as detailed below. The first audit, in October, was generated completely by the NDCS center secretaries according to specifications supplied by the data base manager in Cambridge. The second and subsequent rosters were computer-generated listings based on information received in the previous audit. These listings were simply updated by the secretaries to reflect terminations, new intakes, and error corrections.

Information collected or verified for each child in each audit included birth date, the total number of scheduled in-center hours per week, the total number of days in the center per week, the type of schedule (morning, afternoon or mixed), the child's class assignment, two or three days of absenteeism data, and the termination date if the child was no longer enrolled as of the audit date. With the exception of birthdate and termination date, this information was "as of" a specific date for each audit. Each audit, therefore, represents the center's configuration on a given date. Absenteeism was checked by recording for each child the number of center hours scheduled and the number of hours absent on each of two or three days.* The purpose of this was not to obtain information valid at the child level, but to allow estimation of the overall absentee rate for the class or center as a whole.

Class assignment information was based on a "center structure" form completed by the center secretary with each audit. For each class in the center, the form listed an identification code to be used with the rosters, the principal teacher, a class name used for tracking purposes, an indication of whether major changes had occurred in the child assignments since the previous audit, and an indication whether or not the class was a "target" class--that is, whether it consisted primarily of children who were three to four years old on October 1, 1976. Whereas all children and staff in the center were included in the audits, only target classes were observed in the inspections. The first center structures were collected from the site offices in September 1976 by telephone. Computer-generated center-structure forms were then used for all five of the audits, with successive updates from audit to audit.

Audits of the staff and volunteers in each center were slightly more complex because a given person could have more than one job in the center. Staff and volunteer data collected or verified in each audit included the total number of scheduled in-center hours per week; the total number of days per week; the type of schedule; a code for the primary job; the class assignment for the primary job (if the job was class-related); the scheduled hours per week in the primary job; the code, class assignment, and scheduled hours for the secondary job, if any; absenteeism data; and the termination date if applicable. A volunteer was included in the audit data only if scheduled for 10 or more hours per week in the center. (The inspections, as explained later, did not impose such a criterion.) The system for absentee checking and class assignments was the same as that used for the child audit. The job codes indicated whether the job was an in-class assignment and whether the indivdual was a paid staff person or a volunteer.

---

*Two days were used to check absenteeism in the first three audits; a third day was added in the last two audits.

The foregoing procedures resulted in ten distinct files of audit data: one each for children and for staff for the months of October, December, January, March and April, 1976-1977. The process of computing class- and center-level classroom composition variables from these data included the following steps.

- The total number of children in each class was averaged across all target classes in the center to obtain the mean rostered group size. The total number of scheduled child hours in each class was divided by 40 to obtain the scheduled group size in terms of full-time equivalent (FTE) children.* Nontarget classes were excluded so that maximum congruence with the center inspections would be maintained.

  Similar means were obtained for all non-volunteer staff whose primary or secondary job was a target-class assignment. In the case of schedule data, only the hours which scheduled for class-related jobs were summed. The number of scheduled FTE staff were obtained by dividing the total hours by 40.

- The process for obtaining information on the staff was repeated, this time including volunteer data in the computations to produce information regarding caregivers as a whole. Throughout the remainder of this paper, the

---

*Forty hours was used as the constant for computing FTE's not only because it is traditonal but also because it conforms with actual data. Coelen, Glantz and Calore (1978) found that the number of hours spent in the center by full-time children was 41.8, 40.4, and 38.8 for three-, four- and five-year olds, respectively.

42

term caregiver includes both paid staff and unpaid volunteers. The term staff, which has been used rather loosely heretofore, will henceforth refer only to paid center staff (regardless of whether payment was made by the center or by a third party.)

- For both roster and schedule variables, the number of staff and the number of caregivers were each divided by the group size to obtain staff/child and caregiver/child ratios.

- In order to adjust these variables for absenteeism, child and caregiver absentee rates were computed. (The staff rate was assumed to be the same as the caregiver rate.) These were equal to the total number of hours all target-class children (and caregivers, respectively) were actually present on absentee-check days divided by the aggregate number of hours they were scheduled to be present.* Unadjusted child and adult counts and FTE's were multiplied by the appropriate absentee rates to produce absentee-adjusted variables, which were in turn divided as above to produce absentee-adjusted ratios.

- Because Phase 2 analyses had indicated that logarithmic transformation improved the distributions of ratio and group size variables (such that they more closely met the statistical assumptions underlying the analyses and also significantly affected the correlations obtained), all of the unadjusted and absentee-adjusted variables defined above were computed also in the form of logarithms with a base of ten.**

- It is apparent that a large number of classroom composition variables were generated by these computations. By way of summary it may be noted that the basic set of variables included group size, the number adults with and without volunteers, and adult/child ratio with and without volunteers. This basic set was then obtained in rostered and scheduled versions, with and without absenteeism adjustment, and with and without log transformation.

---

* The resultant quotient is actually a "presence rate" but "absentee rate" is a more familiar term and perhaps less confusing if not taken literally as to its construction.

**A discussion of the use of log-transformed classroom composition variables concludes Section 2.

- All of these variables were constructed at the center level for each of the five audits. The five time points were then averaged, yielding a single Phase III measure for each composition variable.

- The center-level full-year variables thus obtained were used in predicting effects measured by the by the cognitive tests, which also had been aggregated to the center level and which involved May-versus-October change scores. NDCS analyses of the Child-Focused and Adult-Focused Interactions, however, were conducted at the class and teacher levels and only within, not across, the October and May observation periods. It was necessary, therefore, to generate a class-level classroom composition file from a single audit for use with these measures. The April audit was chosen, for use with the May interactions, because the year-end data were presumed to be slightly more reliable than those obtained at the beginning of Phase III. It was not deemed necessary to replicate the May analyses with October data.

Center Inspections of Classroom Composition

Tuesday through Friday of the week following the audit date were used to observe the composition of target classes in each center.* (Recall that target classes were those consisting primarily of children who were three to four years old on October 1, 1976.) On Tuesday and Thursday, each target class was observed once during the hours of 1:00, 2:00, 3:00 and 4:00 p.m. On Wednesday and Friday, each was observed and during the 9:00, 10:00, 11:00 a.m. and 12:00 p.m. hours.*

---

*For the April inspection, the week preceding the audit was used. The decision not to observe on Monday was a matter of scheduling convenience and was not based on substantive issues.

At ten minutes after the beginning of each hour, the center secretary began to observe the first of the target classes in her center, using the Classroom Observation Form shown in Figure 1. She observed for a period of four minutes, recording the number of children present from the target class; the number of children from other classes; the general activity of the children; the total number of adults present; and separate counts of adults classified as bystanders, as volunteers or specialists, and as center staff or substitutes. Bystanders were defined as adults who were not involved with or responsible for the children but were present during the observation. Typically, this was a parent waiting to pick up a child, but could also have been another staff person who happened by while the counts were being made. Volunteers were defined as unpaid caregivers. Specialists were defined as persons who, though usually paid, came to the center only occasionally. Center staff included all regularly scheduled paid adults, even those not paid by or through the center. Persons substituting for center staff were included in the count of such staff. All counts were based on children and adults who were present at the end of the third minute of observation.

The center secretary repeated this four-minute procedure until all target classes in the center were observed. In some centers a six-minute interval occurred between observations so that observations began at ten-minute

---

*Observations were not conducted during other hours because Phase 2 analyses had shown classroom composition to be largely a function of arrival and departure activities during these times. Although arrival and departure may be significant events for the child, center differences in composition resulting from observations conducted at a given moment reflect almost entirely differences in arrival or departure schedules, and not differences in caregiving procedures.

FIGURE 1

National Day Care Study

Abt Associates Inc.

L. CLASSROOM OBSERVATION FORM

OMB #________
Expires:________

OFFICE USE ONLY ☐☐☐☐☐☐☐ 01–07

Center______________________________

Site______________________________

Classroom Identification:

Space Code: ☐☐ Lead Teacher ________________

Class Identifier: ________________________

Date: mo. ☐☐ 08/09 day ☐☐ 10/11 yr. ☐☐ 12/13 ________ Day of Week ☐☐ 14/15

Time of Day: From ☐☐ 16/17 : ☐☐ 18/19 to ☐☐ 20/21 : ☐☐ 22/23

A. PERSONS PRESENT AND GENERAL CHILD ACTIVITY

(1) Number of children from this class .... ☐☐ 24/25

(2) Number of children from other classes .... ☐☐ 26/27
From whose classes? ______________________________

(3) Number of bystanders and other adults not involved with or responsible for children in this observation ☐☐ 28/29

(4) Number of volunteers, specialists, etc. (not including bystanders) .... ☐☐ 30/31

(5) Number of center staff and substitutes (not including bystanders) .... ☐☐ 32/33

(6) Total number of adults present at end of the fourth minute .... ☐☐ 34/35

(7) If one child activity code generally describes what is happening, what is it? (Leave blank if more than one code is absolutely necessary.) .... ☐☐ 36/37

B. ACTIVITY BY GROUPS

Describe how children in this observation are grouped. In general there must be at least three children together to make a group.

CHILD ACTIVITY CODES

- 01 Arrival
- 02 Indoor Free Play
- 03 Outdoor Free Play
- 04 Planned Group Activity
- 05 Snack
- 06 Meal
- 07 Nap
- 08 Individual Activity
- 09 Transition
- 10 Departure
- 11 Other (Describe)

	NUMBER CHILDREN	CHILD ACTIV.	NUMBER ADULTS	DESCRIPTION OR COMMENT
GROUP 1:	☐☐ 38/39	☐☐ 40/41	☐☐ 42/43	____________
GROUP 2:	☐☐ 44/45	☐☐ 46/47	☐☐ 48/49	____________
GROUP 3:	☐☐ 50/51	☐☐ 52/53	☐☐ 54/55	____________
GROUP 4:	☐☐ 56/57	☐☐ 58/59	☐☐ 60/61	____________
GROUP 5:	☐☐ 62/63	☐☐ 64/65	☐☐ 66/67	____________
GROUP 6:	☐☐ 68/69	☐☐ 70/71	☐☐ 72/73	____________

Adults not involved with these groups (do not include bystanders) .... ☐☐ 74/75

Other comments: ____ 2 8 1 0 1
76/77/78/79/80

____

____

intervals. All observations for each hour were completed within the hour, most within 30 minutes.

The foregoing procedure produced five distinct files of observed classroom composition data for Phase III: one each for the months of October, December, January, March, and April, 1976-1977. As mentioned earlier, the data in each file were collected within one week of the corresponding center audit. Observed composition variables were constructed from these data in the following manner.

- Bystander adults were excluded from all computations.
- Group size consisted of the sum of children present from the target class and children present who were from other classes. The observations thus reflect any merging of classes which took place, whether for lunch, nap, or some common activity.*
- Adult counts likewise included both adults specifically assigned to the target class and other non-bystander adults present during the observation. As in the audit computations, separate variables were constructed for the number of staff and for the number of caregivers.
- A staff/child ratio was computed for each observation by dividing the number of staff by the group size. A corresponding caregiver/child ratio was also obtained.
- Log-transformed group size, staff and caregiver counts, and ratios were also obtained for each observation.

---

* In most cases the "group" in "group size" consisted of all the children in a classroom. However, in a few centers organized in an open classroom pattern, there were several clusters of children and caregivers in a single large space. The NDCS treated each of these clusters as a separate group by focusing on a specific teacher and counting all the children and adults in her cluster.

- In both Phase II and Phase III, differences were found in ratios computed at different hours of the day. Some hours, however, were homogeneous--both in terms of ratio and in terms of the child activity typical of the period. It was found that the day consisted of the following segments in almost all centers. In general, between 7:00 and 9:00 children arrive and have breakfast; 9:00 to 12:00 is devoted to planned group activities and/or indoor free play; lunch occurs around 12:00 noon; nap follows between 1:00 and 3:00 p.m.; free play or individual activities take up the remainder of the afternoon; and children begin to leave (often in a trickle) around 3:30 or 4:00 p.m. until all have gone, shortly after 5:00 p.m.*

- It was therefore necessary, in aggregating observations to the class and center levels, to construct composition variables specific to these segments of the day. Furthermore, Phase 2 analyses indicated that the morning segment was most relevant to the issue of classroom composition effects on children both because of the kinds of activities occurring then and because measurement accuracy was greatest for that segment. The observed composition variables used in Phase III NDCS effects analyses, therefore, (and in all analyses reported in this paper) were constructed by averaging all observations made between 9:00 and 11:59 a.m., and only those observations.

- As in the case of the audit-derived variables, observed composition was computed at the center level for each of the five months, then averaged across months to obtain a single Phase III measure for each variable. The resultant file was used in predicting October-May cognitive change scores. To analyze the May Child-Focus and Adult-Focus Interactions, a separate class-level file was constructed from the April inspection data.

---

*Further information concerning fluctuations in ratio and group size during the course of the day and the year may be found in Chapter 2 of the Phase II Research Report. Variance components and generalizability of the Phase III classroom composition variables is discussed in a separate technical paper by Singer, Affholter, and Goodrich (1978).

Because the merging of day care classes profoundly affected the analyses reported in Section 3, it is worth taking a moment to mention the limitations of NDCS data concerning this phenomenon. Although it was possible to indentify the occurrence of merging by comparing the count of children from the target class to the count of children from other classes, no information was collected as to the reason why two (or more) classes had merged. Clearly, however, the reason can be as important as the frequency of merging in a regulatory context. In particular, it might be necessary to distinguish between merging which is planned in advance as part of a scheduled activity and merging which occurs haphazardly or for the convenience of the staff (rather than for the experience of the children). This is not to suggest that all planned merging is acceptable, regardless of the resultant group size, nor that impromptu merging is always to be avoided, but only that regulatory guidelines might have to deal with such a distinction. Insofar as NDCS data are incomplete in this respect, study analyses cannot inform this aspect of the guidelines.

Merging is a phenomenon which may be planned in advance but often is not. In the five Phase III rounds of inspections, the proportion of NDCS target classes which were observed to merge somewhat regularly in the morning varied from 16 to 28 percent. (Many centers merged classes during lunch or nap, but these times of the day were not included in any of the observations analyzed for this paper.) Moreover, not only the number, but the identity of centers which merged classes varied during the course of the year. Although 24 centers were observed to merge classes in at least one of the inspection rounds, only ten did so in three or more rounds and only six did so in all five rounds. Twelve centers, on the other hand, merged classes only once in five rounds of observations spread across the year. (The number of centers which merged classes in any given round varied from a low of 10 in October and January to a high of 16 in April.*)

In summary, the tendency to merge classes is a stable center characteristic only in some centers. In the majority of centers which merge classes at all, it appears to be a relatively unpredictable event. Even in centers where merging does occur regularly throughout the year, it might not always be planned in the sense that it can be indicated in center schedules prepared in advance.

Concerning the Use of Staff/Child, Child/Staff, and Log-Transformed Ratios

Throughout Phase II and Phase III, NDCS reports have expressed ratio in terms of the number of staff available to a group of children: a staff/child ratio. This expression of ratio is useful because it is equivalent to the "fraction of potential staff time" available to each child, and is also the only ratio that can be computed when no staff are present. However, this use of a decimal fraction is less meaningful at a glance than the fully expressed numerical ratio. A staff/child ratio of 1:8 is more immediatly comprehensible than the decimal equivalent of .125. Tables in NDCS reports, therefore, have sometimes reported mean ratio both as a decimal fraction and as a proper fraction with unit numerator (e.g., .15 and 1:6.7).

---

* The April increase may be climate related, since it was almost entirely an Atlanta phenomenon. One might speculate that during the Spring, centers in warm climates have more outdoor play in the morning, and thus more merging as well.

The fact that 1:6.7 (or .15) appears to represent the same quantity as 6.7:1 might seem to suggest that there is no difference between staff/child ratio and child/staff ratio. In fact, however, the two are distinct variables which are not interchangeable. Each is the inverse of the other; they are not linear transformations of one another. For example, the average staff/child ratio of two classes whose separate ratios are 1:5 and 1:10 is (.20+.10)/2 = .15, or 1:6.7. The average child/staff ratio, however, is (5+10)/2, or 7.5:1. If one were to compute a staff/child ratio and a child/staff ratio for each center and conduct statistical analyses of each variable, the results would not and could not be numerically identical, even though they might lead to the same conclusions.* For mathematical reasons alone, therefore, it is important to maintain a distinction between these two variables, and tables which omit the preceding "1:" in NDCS reports should not be mistakenly interpreted as reporting analyses of child/staff ratio.

The conceptual difference between the two forms of computing ratio, however, is somewhat elusive. Staff/child ratio, being the "fraction of caregiver time " potentially available to each child, is more directly related to staff resource allocation and cost per child than is child/staff ratio. Child/staff ratio, on the other hand, is more directly related to labor productivity and, under a given fee schedule, gross center revenue. Nonetheless, the two expressions are merely different scales of the same reality. The distinctions cited can be obliterated by using yet another scale: a simple logarithmic transformation. The logarithm of a child/staff ratio is identical to the negative logarithim of the corresponding staff/child ratio, and therefore the two behave identically in all statistical respects. In the example above, the logs of 1:5 and 1:10 are -.7 and -1, whereas the logs of 5:1 and 10:1 are +.7 and +1. Except for the sign, therefore, the average log-transformed staff/ child ratio is the same as the average log-transformed child/staff ratio.

---

*It may be noted, in fact, that just such parallel analyses were conducted for the Phase II Research Report. The results were indeed completely consistent, leading to the same conclusions in both cases.

Other advantages of log-transformed ratio arise from its equivalence with the difference between the log of the numerator and the log of the denominator. First, the mean of a number of log ratios is equal to the mean of the logs of the numerators minus the mean of the logs of the denominators. This is not true of standard ratios--the mean of several staff/child ratios is not equal to the mean number of staff divided by the mean number of children. Therefore it can be difficult with ratios measured in traditional ways to determine whether a difference between mean ratios is due to a difference in the mean number of staff or to a difference in mean group size. Under a log transformation, however, this difficulty does not arise.

Second, the collinearity among log ratio, log group size, and the log of the number of caregivers simplifies the regression model in that only two of the three variables need be entered. Regression coefficients for log ratio and log group size can be used to compute the regression coefficients for log caregivers. This cannot be done if unlogged variables are used because, as mentioned above, the mean of a ratio does not equal the ratio of the means.

Finally, the use of a log transformation reduces or eliminates the sometimes considerable skew and kurtosis shown in the distribution of untransformed ratio and group size, bringing the data closer to conformity with the assumptions underlying parametric statistical analyses. This is especially helpful in computing correlations, which are known to be more affected by distributional distortion than are tests of mean differences.

Because of these advantages, log transformations of composition variables were used where possible for analyses reported in this paper, although table means are shown in terms of their antilogs. (An antilog, simply stated, is the untransformed number corresponding to a logarithm. If the log of 100 is 2, for example, then the antilog of 2 is 100.)

CHAPTER THREE: COMPARISONS AMONG ROSTERED, SCHEDULED AND OBSERVED CLASSROOM COMPOSITION

This section presents the results of analyses comparing rostered, scheduled and observed classroom composition. These analyses included not only direct comparisons of each set with the other but also investigations of the effects of whether a center tended to merge classes, of adjusting vs. not adjusting audit variables for absenteeism, of including vs. excluding volunteers, and of transforming vs. not transforming the variables to a log scale.

In order to simplify the text and tables which follow, means of log-transformed variables are given in terms of their antilogs. Care should be taken, however, to keep in mind whether log-transformed or untransformed variables are under discussion. This is most easily done by reference to the table in question. It should be mentioned that the antilog group sizes and numbers of caregivers cited in the text and tables are lower than the simple means of untransformed "standard" group size variables.*

Table 1 presents direct comparisons among rostered, scheduled, and inspected variables for the entire sample of

---

*This is why means reported in this paper differ slightly from corresponding means reported in Volume 1 (Children at the Center); "standard" classroom composition variables were analyzed for Volume 1.

57 centers.* Note that mean rostered group size and mean scheduled group size are the same (17.8) and both are larger than the mean inspected group size (15.5). (It will be seen later that absenteeism largely accounts for this difference.) The mean rostered ratio, however, is smaller than the mean scheduled ratio (1:5.3 vs. 1:7.1) which, in turn, is the same as the inspected ratio. This implies that there is a difference between the rostered and the scheduled number of caregivers, which, indeed, is the case. These means are 3.4 and 2.5, respectively.

The Influence of Merging on Classroom Composition Measurements

The results in Table 1 are potentially misleading because of the practice, followed in many centers, of merging classes at various times of the day. Quite obviously, the observed group size in two classes which are merged would be twice that of the scheduled group size of either, assuming that the same number of children are scheduled for each class. The observed ratio, however, could well be the same if the caregivers assigned to both classes have all remained to supervise the larger group.

Specific observation information regarding the number of children present from the class being observed and the number present from other classes made it possible to determine whether merging had occurred.

In an attempt to distinguish classes which rarely merge from those in which merging is common, several separate

---

*Table 1 variables are log transformed and include volunteers; roster and schedule variables have not been adjusted for absenteeism.

Table 1

Comparison of Log-Transformed Rostered, Scheduled, and Inspected Classroom Composition (N=57 Centers)

MEANS[a]

	Rostered	Scheduled	Inspected
Log Group Size	17.8	17.8	15.5
Log Number of Caregivers	3.4	2.5	2.2
Log Caregiver/ Child Ratio	1:5.3	1:7.1	1:7.1

CORRELATIONS

	Log Group Size			Log No. Caregivers			Log Ratio		
	Rost.	Sched.	Insp.	Rost.	Sched.	Insp.	Rost.	Sched.	Insp.
Rost.	-			-			-		
Sched	.96	-		.87	-		.85	-	
Insp.	.67	.69	-	.56	.74	-	.52	.70	-

---

[a] Means are presented in terms of their antilogs to aid interpretation. Rostered and scheduled group size and number of caregivers both differ significantly from inspected group size and number of caregivers (p<.005). Rostered ratio differs significantly from the inspections (p<.001), but scheduled ratio does not.

decision criteria were applied to the six morning observations conducted in April 1977 (T4). These criteria, which resulted in the same decision in virtually every case, defined a class as merging at least occasionally if

- the average number of children from other classes (across the six observations) was 10 or more.
- the average proportion of children from other classes, relative to the total group size, was 30% or more.
- at least two of the six observations reported five or more children present from other classes.
- at least two of the observations reported that more than one-third of the children present were from other classes.

Using these criteria in concert, 96 of the 133 target classes in April were identified as rarely, if ever, merging and 37 were identified as classes which merge occasionally or more often.*

It must be stressed again, as it was in Section 2.0, that merging is a practice which can be planned in advance as a regular part of the schedule or may happen at random times, at the convenience of the staff or during isolated non-recurring events. From a monitoring standpoint, therefore, it is a practice requiring flexible guidelines for interpreting the diverse reasons centers have (or don't have) for merging classes. Data from the NDCS cannot inform these guidelines, because insufficient information is available concerning the reasons for and the regularity of the merging which was observed. All that can be reported here are the differences between merging and nonmerging centers regarding observed vs audited measures of classroom composition. It must be pointed out, however, that merging would not have been detected, nor its importance realized, had schedules been relied upon as the sole composition measure, and observations not conducted.

---

*In analyses which follow, one of the 37 merging classes was eliminated becasue it was a large single-classroom unstructured center which functioned as a statistical outlier in the regressions. The antilog of its observed group size was 10.7 and of its scheduled group size 52.0. Although statistically an outlier, this one example is strong evidence of the impracticality of monitoring open-classroom centers with audits of schedules.

Table 2 presents comparisons between April 1977 scheduled and inspected group size and ratio, but separately for the 96 nonmerging target classes and the 36 merging target classes. The results are striking. Scheduled group size is only slightly higher than observed group size among classes that do not merge, and this difference, as will be seen, is entirely accounted for by absenteeism. Among merging classes, however, scheduled group size is lower than observed group size by more than six children--and adjustment for absenteeism enlarges this difference even further. The correlation between scheduled and observed group size is not affected by merging, the difference between .85 and .72 not being statistically significant. Exactly the same pattern of results can be noted for the number of caregivers in these 132 classes.

Not surprisingly caregiver/child ratio is relatively unaffected by merging: there is no difference between scheduled and observed ratio for either the nonmerging or the merging sample. This merely indicates that when the children in two (or more) classes merged, the caregivers in those classes tended to merge as well.

Table 2

Comparison of April 1977 Log-Transformed Scheduled and Inspected Classroom Composition: 96 Nonmerging Classes Versus 36 Merging Classes[a]

96 Classes Which Rarely Merge

	Scheduled	Inspected	r	t	p_t
Log Group Size	17.0	15.0	.85	5.9	<.001
Log No. Caregivers	2.2	2.0	.81	3.0	.003
Log Caregiver/Child Ratio	1:7.8	1:7.6	.64	-1.0	.322

36 Classes Which Sometimes Merge

	Scheduled	Inspected	r	t	p_t
Log Group Size	13.3	19.7	.72	-8.3	<.001
Log No. Caregivers	1.7	2.4	.64	-5.2	.001
Log Caregiver/Child Ratio	1:7.9	1:8.1	.69	0.5	.637

---

[a] Means are presented in terms of their antilogs to aid interpretation.

The Effect of Absenteeism on Scheduled Composition

In the introductory section to this paper it was mentioned that one advantage of observations is their inherent adjustment for absenteeism. In the present section, moreover, it was stated that the difference between observed and scheduled group size among nonmerging centers disappears when the schedules are adjusted for absenteeism, but that this adjustment only enlarges the difference for merging classes.

The results to substantiate these statements are presented in Table 3, which compares observed composition variables with absentee-adjusted schedule variables, separately for nonmerging and merging casses. The format of Table 3 is the same as that for Table 2 in order to facilitate the further comparison between unadjusted and adjusted scheduled classroom composition.

In nonmerging classes, observed group size corresponds quite well to absentee-adjusted scheduled group size--the means are virtually identical and the correlation between the two is .83. Observed and scheduled ratio do not correspond so well when measured at the class level--the means are identical, but the correlation is only .58. This, however, was one reason why the recommendation was made to monitor ratio at center level: the correlation between center-level observed ratio and scheduled ratio for those centers which rarely merged classed was .76. For centers which do not usually merge classes, therefore, observations are reasonably congruent with schedules when measuring group size at the class level and rztio at the center level.

Among merging classes, the differences between observations and adjusted schedules are dramatic, but only for group size (a difference of eight children) and for the

Table 3

Comparison of Inspected Classroom Composition with Absentee-Adjusted Scheduled Composition: April 1977 Data for 96 Nonmerging Classes Versus 36 Merging Classes[a]

96 Classes Which Rarely Merge

	Scheduled	Inspected	r	t	p_t
Log Group Size	15.1	15.0	.83	0.2	<.836
Log No. Caregivers	2.0	2.0	.75	-	-
Log Caregiver/Child Ratio	1:7.6	1:7.6	.58	-	-

36 Classes Which Sometimes Merge

	Scheduled	Inspected	r	t	p_t
Log Group Size	11.6	19.7	.71	-11.0	<.001
Log No. Caregivers	1.5	2.4	.68	-6.1	<.001
Log Caregiver/Child Ratio	1:7.6	1:8.1	.77	1.4	.174

---

[a] Means are presented in terms of their antilogs to aid interpretation. The correlations between adjusted and unadjusted scheduled composition variables for both samples range from .93 to .98.

number of caregivers (a difference of nearly one caregiver). As in Table 2, merging is not seen to significantly affect the difference between observed ratio and the adjusted schedule ratio. (The difference between the .58 correlation for nonmerging classes and the .77 correlation for merging classes is not statistically significant.)

Finally, the correlations between adjusted and unadjusted composition variables was extremely high. Not shown in the tables, they ranged from .93 to .98. This merely indicates that the absentee rate, which was sufficiently high to have an appreciable effect on the center mean, did not vary much from center to center. After adjustment, centers remained in virtually the same location relative to one another--thus the correlations were high.

It should be noted that the 36 merging classes in these analyses do not represent centers in which _all_ of the classes merged. At the center-level, for example, 15 of the 56* study centers were identified as having either a majority of classes which merge or a substantial minority in which merging was highly "visible" from the observation data. Although these 15 centers account for 34 of the 36 merging classes, they also account for 16 of the _nonmerging_ classes. On the other hand, only 2 of the 82 classes in the 41 nonmerging centers were classes which tended to merge. Predictably, therefore, _center_-level means for scheduled vs. observed group size (and number of caregivers) were less disparate for merging centers than were the _class_-level means reported for merging classes in Table 3. Curiously, however, the differences in correlations (as a function of merging) were more pronounced at the center level. Instead of group size correlations of .83 and .71 for nonmerging and merging classes, the corresponding center-level correlations were .94 and .39. Complete center-level results corresponding to those in Table 3 are presented in Table 4.

---

*The open-classroom center eliminated from the class-level analyses was likewise eliminated from the center-level analyses.

The Effect of Volunteerism on Caregiver Measurement

One of the more controversial issues regarding federal regulation of day care centers has been whether volunteers should be included in adult counts used to compute ratios. Table 5 rather conclusively shows that there is virtually no difference between classroom composition measures that include volunteers and those which don't. The correlations range from .97 to .99 and the means are nearly identical. The mean number of adults, for example, is only a tenth of a person higher, at most, when volunteers are included. The results in Table 5 are especially striking given the different criteria used for the audits and the inspections regarding volunteers. Audit data were not even collected on volunteers who were scheduled to spend less than ten hours a week in the center. For the inspections, however, any unpaid adult who was observed in the classroom (other than a bystander) was considered a volunteer and counted as such. Despite this more inclusive criterion, the inspection data are as compelling as the audit data in defusing the volunteer issue--at least in classes of three- and four-year-olds. Moreover, similar conclusions were reached by Coelen, Glantz and Calore (1978), who reviewed self-reported audits from a national sample of over 1500 centers.

It should be noted, however, that these conclusions are based on averages across a sample of centers. There are in the United States some centers, such as parent cooperatives, for whom volunteers are the major, if not the only, source of caregivers. Though they may be few in number, these centers would be radically affected by regulations which disallowed the inclusion of volunteers in ratio computations.

Table 4

Comparison of Inspected Classroom Composition with Absentee-Adjusted Scheduled Composition: April 1977 Data for 41 Nonmerging Centers Versus 15 Merging Centers[a]

41 Centers Which Rarely Merge Classes

	Scheduled	Inspected	r	t	p_t
Log Group Size	16.6	15.9	.94	2.1	<.047
Log No. Caregivers	2.5	2.3	.89	2.2	.032
Log Caregiver/Child Ratio	1:6.7	1:6.9	.76	0.8	.417

15 Centers Which Sometimes Merge Classes

	Scheduled	Inspected	r	t	p_t
Log Group Size	11.9	14.9	.39	-2.7	.019
Log No. Caregivers	1.9	1.9	.39	-	-
Log Caregiver/Child Ratio	1:6.5	1:7.7	.59	1.8	.092

---

[a] Means are presented in terms of their antilogs to aid interpretation. One large open-classroom center was eliminated from these analyses as a statistical outlier.

Table 5

Comparison of Classroom Composition Excluding Volunteers ("Staff") and Including Volunteers ("Caregivers"): N=57 Centers[a]

	Number of Staff	Number of Caregivers	Pearson r
Log Rostered	3.3	3.4	.992
Log Scheduled	2.5	2.5	.995
Log Inspected	2.1	2.2	.984

	Staff/Child Ratio	Caregiver/Child Ratio	Pearson r
Log Rostered	1:5.4	1:5.3	.987
Log Scheduled	1:7.2	1:7.1	.992
Log Inspected	1:7.4	1:7.1	.970

---

[a] Means are presented in terms of their antilogs to aid interpretation. A volunteer was included in the rosters and schedules only if scheduled for ten or more hours per week in the center. For the inspections, however, any unpaid adult observed in the classroom was considered to be a volunteer. In addition, the caregiver count for inspections included specialists who, though paid, were only occasionally in the center.

composition. Although the correlation (not shown in Table 6) between each logged variable and its untransformed counterpart ranged from .92 to .97, skew and kurtosis in the unlogged distributions were improved sufficiently to have a substantial impact on correlations with other variables. The correlations among logged composition variables (on the left side of Table 6) are all higher than the corresponding correlations among the untransformed variables, by as much as .15 in one instance. One example of the effect of log transformation on the distributions is a reduction in skew for rostered group size (in terms of the normal or Z distribution) from 1.86 to .71 and a reduction in kurtosis from 4.39 to -.04.

Table 6

Comparison of Intercorrelations With and Without Log Transformation (N=57 Centers)

	LOGGED				UNLOGGED			
		Insp.	Rost.	Sched.		Insp.	Rost.	Sched.
GROUP SIZE	Insp.	-			Insp.	-		
	Rost.	.67	-		Rost.	.52	-	
	Sched.	.69	.96	-	Sched.	.59	.95	-
		Insp.	Rost.	Sched.		Insp.	Rost.	Sched.
	Insp.	-			Insp.	-		
RATIO	Rost.	.52	-		Rost.	.39	-	
	Sched.	.70	.85	-	Sched.	.61	.81	-

CHAPTER FOUR: THE FEASIBILITY OF MONITORING REGULATORY COMPLIANCE WITH OBSERVATIONS

It is clear from the analyses reported here that center-prepared schedules of classroom composition cannot by themselves validly reflect configurations of children and caregivers actually occurring under all conditions. Only observations appear capable of dealing with unusual circumstances (such as unstructured, open-classroom environments in large centers) and unplanned events (such as the extemporaneous merging of two classes). The incidence of merging, moreover, and the magnitude of its impact on composition measurement, strongly suggest a reassessment of the method by which compliance with day care regulations should be monitored.

During Phase II and most of Phase III, there was some concern on the part of NDCS staff that observations were the principal source of ratio and group size information. The reason was that the person-hours devoted to these observations were considerable. Recall, for example, that every six weeks or so during Phase III each target class in each center was observed once each hour for four half-days. It was clear that licensing and other monitoring agencies would never be able to devote this much time to checking compliance. But what if analyses indicated that this level of effort was required to achieve adequate accuracy of measurement?

Part of this concern was resolved by the realization that monitoring compliance might not require periodic observations over the course of the year for the purpose of computing a yearly average. Future regulations might specify that compliance decisions be made each time compliance measurements are made, being based simply on the classroom composition in effect at the time.

Although a single observation can show whether a center is within the legal limits at the moment the counts are made, however, these counts can change from moment to moment and day to day. If no allowance whatsoever is made for such fluctuation, the monitoring process could degenerate to an absurdity: a center might be out of compliance while a teacher went to the bathroom but in compliance when she returned. Future day care regulations, which assuredly would take such fluctuations into account, could do so by specifying, in the monitoring guidelines, the span of time (week, month, or year, for example) to which classroom composition measurements must generalize. The issue in monitoring by means of observations, therefore, concerns the number which must be conducted in order to "smooth-out" short-term fluctuations within this span of time.

In order to investigate this issue, variance components analyses of each of the five Phase III inspections were conducted. These analyses were somewhat different from the analyses reported by Singer, Affholter and Goodrich (1978), because their purpose was different. As in all NDCS analyses, only morning observations were used, because these hours are most relevant to the effects which the NDCS sought to detect. Unlike those analyses, however, observations conducted after 11:00 a.m. were excluded. Inspection of the data had revealed that a number of centers were having lunch during these observations, and it was not desirable to confound this variance with other variance attributable to hourly fluctuations during the morning. Furthermore, only one classroom (rather than two) were sampled in each center, in order to make use of observations from all the centers in the NDCS.*

---

*Including two classes would have forced the exclusion of one-class centers because the variance components computer program required a balanced design. Yet previous analyses had shown variation among classes within a center to be negligible compared to the variation among centers.

The analyses, in sum, were designed to yield point estimates of the following variance components:

- Center/Class (confounded, but probably almost entirely center variation)
- Time of the Morning (9:00 vs. 10:00 a.m.)
- Center x Time Interaction
- Occasion (nested within center x time and obtained by observing centers on one Wednesday and one Friday morning during each round of inspections)

It was decided, however, that the inclusion of centers which merge classes would be inappropriate. The rationale was that the merging of two or more classrooms is a highly visible,* discrete event which could (and should) be dealt with explicitly by the procedural guidelines used in monitoring centers via observation. It seems highly unlikely that such guidelines would recommend that observations proceed as usual when merging is noted, but would more likely recommend that the observer stop at that point and start a different procedure--perhaps an interview with the director concerning the center's practice of merging classes. In order that the variance components analyses conform as closely as possible with procedural guidelines likely to exist in a compliance-monitoring context, centers with merged classes were excluded from each of the five analyses. Finally, standard rather than log-transformed variables were analyzed, here again in order to maximize correspondence with potential guidelines.

Only two numbers from each variance components analysis are relevant to this investigation: the variance due to time and the variance due to occasion. The sum of these two represents the short-term variation cited earlier which must be absorbed as error if observations are used to monitor compliance. These figures are presented in Table 7.

---

*With the director's help, the class structure of a center is easily determined, and an observer can learn what to anticipate regarding the general size of each class to be observed.

The two most important entries in Table 7 are the averages for total short-term variation across the five inspections: 5.32 for group size and .00246 for ratio. The square roots of these entries--2.3 and .05, respectively--are estimates of the standard deviations which could be used to establish a confidence band around the observations of a particular center. Suppose, for example, that regulations limited the maximum group size for a class with two caregivers to 15 and that the observed group size in such a class happened to be 10, or 2.17 standard deviations below the allowable maximum. The likelihood is 99 percent that the class's true group size for that month was within the regulated limit. If the observed group size had been 14, however, or only .43 standard deviations below the limit, then there would only be a 67 percent likelihood that the class was truly in compliance. (The laws of probability, of course, operate in both directions. A class whose observed group size was 16 would appear not to be in compliance, even though there is a 33 percent chance that its true group size that month was 15 or lower.)*

Statistical information, of course, can only help inform a decision about the compliance of a given center; it is clear that such a decision must also take into account

---

*There are problems with such confidence bands however. In particular, they assume that the same standard deviation is appropriate regardless of the group size or ratio observed. This, however, is not the case. For example, large-class centers have a larger standard deviation than small-class centers. The standard deviation for centers with an average group size (across occassions and time) of 19 or more was 2.5, while the standard deviation for centers with an average group size less than 13 was 1.9. See Cronbach, et al., (1972, Chapter 5) for a detailed discussion of this and related issues.

Table 7

Variance Components Point Estimates of Short-Term Variation in Inspected Group Size and Ratio

	Unlogged Group Size					
Source	October	December	January	March	April	Average
Time of Morning	.10	0	0	0	0	.02
Day of Month	5.68	4.54	6.18	5.82	4.31	5.30
Total	5.78	4.54	6.18	5.82	4.31	5.32

	Unlogged Caregiver/Child Ratio					
Source	October	December	January	March	April	Average
Time of Morning	.0000317	0	0	0	0	0
Day of Month	.00526	.00247	.00208	.00165	.000842	.00246
Total	.00529	.00247	.00208	.00165	.000842	.00246

previous knowledge about the center and perceived consequences of decision errors. In forming decisionmaking strategies, it is also appropriate to consider the relevant political realities associated with the consequences of decision errors.

The model which was used for these variance components analyses implies that monitoring a center with observations would involve a single morning visit during which each class would be observed once during the 9:00 a.m. hour and once during the 10:00 a.m. hour. There are several important points that must be made regarding this procedure.

- The exact times at which the observations are conducted is probably less important than the activities ongoing at the time. In particular, children should have already arrived and the day's activities begun; observations should terminate before lunch; and merged classes should be dealt with as a special case. With these exceptions, these results do not suggest limiting observations to specific situations such as planned group activities, free play, etc.

- The compliance of the center as a whole, not that of specific classes, is likely to be at issue, yet observations are necessarily conducted at the class level. Additional computations, therefore, are required--averaging across classes while adjusting for the age range of each (assuming that regulations are age specific). Further discussion of this issue may be found in Day Care in the United States: 1976-1977, by Coelen, Glantz, and Calore, (1978).

- Data used in these analyses are from "target" classes of three- and four-year-olds. It is not known whether the results would differ considerably for younger or older classes.

- The level of accuracy obtainable from a one-morning visit is likely to be higher than estimated by these analyses for three reasons First, these analyses presumed that only one class in the center is observed, whereas in actuality all classes would be. The typical day care center in the NDCS had more than one

> target class, and accuracy is increased when data are averaged across classes. Second, one-classroom centers are likely to show less short-term variation than reported here because at least some of that variation results from children wandering back and forth between various classes. Third, in actual practice, each class could be observed every half hour rather than every hour, since an observation requires only 3-5 minutes. A substantial center x time variance component for group size suggests that this would yield more accurate measurement of group size.* No corresponding component was found for ratio, however, indicating that ratio accuracy would not be increased as a rule. These arguments imply that the standard deviations of 2.3 and .05 for computing group size and ratio confidence intervals are too large, but their revision through further analysis is not possible.

Conclusions

The variance components analyses of center inspections suggest that monitoring compliance by observation is feasible, at least in terms of measurement accuracy. Furthermore, there is reason to believe that the level of accuracy obtained in actual practice would be greater than the level estimated by these analyses.

Inasmuch as one two-hour center visit may be adequate to monitor compliance by inspection, (depending on the span of time to which the estimate must generalize), observations would also seem feasible from a logistical viewpoint. Certainly the personnel demands seem manageable, especially since the collection of large amounts of detailed audit data could be dispensed with. If compliance is to be based on average composition across the year, more frequent than yearly checks may be necessary, but this might be possible through some combination of self-report and spot checks. Center directors could conduct their own observations quarterly, sending the data to the overseeing agency. Agency spot-checks could be used principally to verify the self-reports. Followup inspections would be mandated either if a center was not in compliance or if the spot check differed from the self-report by a specified margin. (In fact, long-term data collection would eventually establish what that margin should be.)

---

*Technically, the variance components design does not permit this conclusion because the 9:00 and 10:00 observatins were assumed to sample the complete population of available morning hours. Inspection of the data, however, suggests that 9:30 and 10:30 would not be trivial redundancies of the activities ongoing at 9:00 and 10:00. The NDCS observation does, however, do not permit the use of an every-half-hour variance comonent design, so the increase in accuracy cannot be estimated.

While this arrangement sounds quite similar to one in which rosters and schedules are reported, it differs from an audit system in two important respects. First, the volume of data generated by an observation system is considerably smaller, principally because records need not be maintained in detail at the child and caregiver level. Second, the demands placed on a center director by an observation system are negligible compared to the resources required to maintain a continuously updated audit system. In order to conduct observations, a director need only know what classes there are in the center. And because an observation is a snapshot, the director need not know (nor keep records of) how many children entered or terminated since the last measure was taken and what their schedules were. In fact, if agency spot checks were the only monitoring activity required by regulations, there would be virtually no demand on center resources. Apart from the cost factor itself, lowering the expense forced upon the center lowers the perceived intrusiveness of federal regulation and lowers the disruption in day-to-day operation which is inevitable if detailed recordkeeping is mandated.

These considerations speak to yet a third issue independent of measurement and logistical feasibility: namely, political feasibility. Besides the question of federal intrusiveness, the simpler the monitoring system the

Lest the foregoing discussion appear too strongly to favor a system of inspections for monitoring compliance, let it be clearly stated that no system is without flaw and equally acceptable to all. The principal conclusion of this paper is not that inspections must surely replace audits, but merely that inspections have distinct advantages--chiefly their superior validity--and are also more practical than might have once been thought. They should be given serious consideration in formulating monitoring mechanisms which accompany any forthcoming federal day care regulations.

A Psychometric Analysis of the National Day Care Study Phase III Child Test Battery

William L. Bache III

77

TABLE OF CONTENTS

Page

CHAPTER ONE: OVERVIEW 53

Description of the Instruments 54
Summary of Findings 55

CHAPTER TWO: PRESCHOOL INVENTORY 63

CHAPTER THREE: PEABODY PICTURE VOCABULARY TEST 69

CHAPTER FOUR: THE SRI TEST MOTOR SKILLS 72

CHAPTER FIVE: PUPIL OBSERVATION CHECKLIST 73

CHAPTER SIX: CONCLUSIONS 78

NOTE 79

REFERENCES 81

APPENDIX A 83

CHAPTER ONE: OVERVIEW

The Phase III child test battery was identical for both the Fall 1976 (T3) and the Spring 1977 (T4) administrations. The two-day testing schedule was as follows:

Day 1: 32-Item Preschool Inventory (PSI)
SRI Fine Motor Test (FMOTOR)
Pupil Observation Checklist 1 (POCL1)

Day 2: 90-Item Revised Peabody Picture Vocabulary Test (PPVT)
SRI Gross Motor Test (GMOTOR)
Pupil Observation Checklist 2 (POCL2)

The Stanford Research Institute (SRI) administered all tests in the National Day Care Study.

Sample sizes tested at each time point, including the number of children in each six-month age interval, are presented in Table 1. Only children who were tested at T3 were retested at T4, resulting in the smaller _n_ for the T4 sample.

Table 1

Numbers of Children Tested at T3 and T4

	T3	T4
Total Number Tested	1463	1113
By Age Group:		
37 - 42	314	--
43 - 48	359	213
49 - 54	389	251
55 - 60	321	308
61 - 66	--	289
Total	1383	1061

Throughout this paper, "total sample" refers to children 37-60 months old at T3 or 43-66 months old at T4. Children outside this age range were excluded from the psychometric analyses.

Description of the Instruments

The Preschool Inventory (PSI) is a test of general knowledge and cognitive skills. Developed by Bettye Caldwell for the Educational Testing Service (ETS, 1970), it has previously been used in the Head Start Longitudinal Study and Head Start Planned Variation Study. The test was not designed as a measure of general intelligence, but as a measure of school readiness. The version used during Phase III was the 32-item form culled from the 64-item PSI by Stanford Research Instituite in 1970-71 for use in the third year of the Head Start Planned Variation study. It was also used in the national evaluation of Home Start. Phase II analyses comparing the shorter and longer forms indicated that very little psychometric precision would be sacrificed by using the 32-item test, while sufficient testing time would be saved to allow the addition of the PPVT to the battery. In the Head Start Longitudinal Study, the 64-item PSI given at age four correlated .59 with third grade reading and math achievement scores and .64 with the Ravens Colored Progressive Matrices (Shipman, McKee and Bridgeman, 1976).

The PPVT is a measure of receptive language which has been used extensively in a variety of research applications. The version used in the NDCS differs from the 1959 edition developed by Lloyd Dunn in two respects. First, SRI used the revised plates developed by the Educational Testing Service for the Head Start Longitudinal Study (Meissner, Shipman, and Gilbert, 1972). The ETS revision

80

was designed to avoid potential cultural bias and improve the relevance of the plates for black children by increasing the number of black persons in the illustrations and by diversifying the roles they represent. (The original PPVT contained only two black figures: a Pullman porter and an African native.)

Secondly, the test was shortened from 150 items to 90. When SRI pretested the first 60 items of the PPVT, no child missed any of the first ten items, while on the other hand some children who answered a large number of items seemed to tire of the test. It was therefore decided to begin the PPVT with item 11. Pretesting also showed that only a few children answered all 60 items without failing 6 of 8 consecutive items--the criterion for terminating the test. Test booklets containing items 11 through 100 (in addition to practive items) were therefore prepared for the NDCS study. Items 101-150 of the ETS revised version were not used.

The SRI Fine and Gross Motor Tests consist of items which are common to many standardized tests of motor skills such as the McCarthy Scales of Children's Abilities and the Denver Developmental Screening Test:

FINE MOTOR

1) Copy a circle
2) Copy a plus sign
3) Draw a person (six body parts)
4) Build a tower of eight blocks
5) Build a bridge with blocks

GROSS MOTOR

1) Balance on one foot for ten seconds
2) Jump in place
3) Jump over width of sheet of paper
4) Take two hops on one foot

5) Walk forward heel-to-toe four steps
6) Walk backward heel-to-toe four steps
7) Catch bounced ball three times

SRI pretested these items in the summer of 1976 and it was expected that separate fine and gross motor scores would be obtained from the NDCS Phase III data. Ceiling effects and inadequate reliability, however, dictated that the 12 items be treated as a single SRI Test of Motor Skills (MOTOR), whose score was simply the number of items passed.

The POCL1 and the POCL2 are the same instrument: a set of nine five-point ratings completed by the tester at the end of each session. Analyses in the National Home Start Evaluation and in Phase II of the NDCS consistently replicated two POCL components traditionally labeled Task Orientation and Sociability and consisting of the following items:*

TASK ORIENTATION

1) Resistive - Cooperative
2) Involved - Indifferent
3) Defensive - Agreeable
4) Gives Up - Keeps Trying
5) Attentive - Inattentive

---

*Items were reversed as necessary so that a rating of "1" would reflect low task orientation and low sociability.

SOCIABILITY

1) Shy - Sociable
2) Outgoing - Withdrawn
3) Active - Passive
4) Quiet - Talkative

Summary of Findings

Subsequent sections of this chapter consist of psychometric results presented separately for each test in the battery. These results are summarized below.

- The reliability of the PSI and the PPVT was very good, their T3 alpha coefficients for the total sample being .84 and .96, respectively. T3-T4 test-retest correlations were also quite high: .77 and.80. There appeared to be no floor effect and only a slight ceiling effect for the PSI, but the PPVT exhibited a rather clear-cut floor effect.

- As expected, the PSI and the PPVT were highly correlated (T3 *r* = .74), though somewhat less so when age was partially out (T3 partial *r* = .64). Since only 41 percent of the variance in either test is predictable from the other when age is controlled, there is good justification for using them as separate dependent variables.

- MOTOR change scores appear to be invalid for the NDCS sample for two reasons. First, a pronounced ceiling effect with children older than 54 months would restrict the computation of such scores to children who were younger than 49 months at T3. Second, even with these younger children, the T3-T4 correlation (r = .52) so closely approximates the scale's reliability (alpha = .59 at T3, .50 at T4) that the reliability of the change scores themselves would be practically nil. Individual differences in MOTOR change scores would be almost entirely attributable to measurement error and not to true change in the level of motor skill.

- POCL Task Orientation change scores are invalid because of an exaggerated positive response bias on the part of the raters. Over 40 percent of the children rated received the maximum possible score at T4. Sociability ratings did not suffer so dramatically from response bias, but exhibited other problems which likewise invalidate their use in change scores. Analysis of Phase II data* revealed that short-term rate-rerate reliability involving different raters and different testing sessions (r = .44) was nearly equal to five-month rate-rerate correlations (r = .37 - .42). Thus, as with the MOTOR scale, change scores would reflect measurement error almost exclusively, rather than true change in sociability.

- Because centers were not tested in the same order at T4 as they were at T3, considerable variation occurred in the exact interval between Fall and Spring testing. Center averages of intertest interval ranged from 190 to 223 days--a difference of one month. Given the sensitivity of the PSI and the PPVT to age, intertest interval was incorporated as an adjustment in the computation of PSI and PPVT generalized gain scores (see Goodrich and Singer, 1979).

Tables 2, 3, 4, and 5 summarize the analyses reported in this chapter. Table 2 presents means and standard deviations of the test scores. Table 3 reports reliability and T3-T4 stability coefficients, and Tables 4 and 5 show the intertest correlations for each time point, with and without controlling for age. The remainder of this chapter discusses the psychometric properties of each test in greater detail.

---

*Replication of this analysis with Phase III data was not possible because, at both T3 and T4, each child was tested by the same person during both test sessions. An accurate estimate of POCL unreliability (including *all* sources of error) is therefore not obtainable from Phase III ratings.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR THE PHASE III TEST BATTERY[1]

T3 AGE		T3PSI		T3PPVT		T3MOTOR		T3POCL1T		T3POCL1S	
(MONTH)	N	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*
37-42	282-304	9.8	*4.1*	12.8	*9.6*	5.0	*1.6*	18.7	*4.6*	12.6	*4.4*
43-48	330-349	12.6	*5.1*	19.4	*12.4*	6.4	*1.8*	19.6	*4.5*	13.0	*4.2*
49-54	372-381	15.5	*5.2*	25.4	*12.3*	7.6	*1.8*	20.8	*4.1*	13.1	*4.2*
55-60	305-318	18.8	*5.3*	30.8	*12.4*	8.5	*1.8*	21.5	*3.9*	13.4	*4.1*
37-60	1297-1352	14.3	*6.0*	22.4	*13.5*	6.9	*2.2*	20.2	*4.4*	13.0	*4.2*

T4AGE		T4PSI		T4PPVT		T3MOTOR		T3POCL1T		T4PQCL1S	
(MONTHS)	N	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*	MEAN	*SD*
43-48	203- 217	14.4	*4.6*	20.1	*11.7*	6.7	*1.6*	21.3	*4.1*	13.2	*4.8*
49-54	235- 244	17.0	*5.4*	26.7	*12.0*	7.9	*1.7*	22.0	*3.4*	13.4	*4.6*
55-60	286- 307	19.8	*5.1*	32.3	*13.1*	8.8	*1.6*	23.0	*2.8*	13.8	*4.6*
61-66	272- 287	22.4	*4.7*	36.7	*11.3*	9.4	*1.7*	23.3	*2.7*	14.0	*4.3*
43-66	996-1054	19.8	*5.8*	29.7	*13.5*	8.3	*1.9*	22.5	*3.3*	13.7	*4.5*

[1] Minimum-maximum possible scores and variable abbreviations:

PSI	(Preschool Inventory)	0-32
PPVT	(Peabody Picture Vocabulary Test)	1-90
MOTOR	(Overall Gross and Fine Motor Score)	0-12
POCL1T	(Day 1 Pupil Observation Checklist: Task Orientation)	5-25
POCL1S	(Day 1 Pupil Observation Checklist: Sociability)	4-20

TABLE 3

RELIABILITY COEFFICIENTS FOR THE PHASE III TEST BATTERY[1]

T3 ALPHA COEFFICIENTS

T3AGE	N	T3PSI	T3PPVT	T3MOTOR	T3POCL1T	T3POCL1S
37-42	282-302	.71	.93	.48	.89	.91
43-48	330-345	.78	.95	.60	.91	.90
49-54	372-383	.78	.94	.58	.91	.89
55-60	305-317	.80	.95	.58	.91	.87
37-60	1297-1344	.84	.96	.70	.91	.89

T4 ALPHA COEFFICIENTS

T4AGE	N	T4PSI	T4PPVT	T4MOTOR	T4POCL1T	T4POCL1S
43-48	200- 215	.71	.94	.43	.87	.92
49-54	235- 242	.80	.94	.52	.84	.91
55-60	286- 307	.79	.95	.46	.86	.93
61-66	272- 285	.78	.94	.50	.87	.89
43-66	996-1045	.83	.95	.60	.87	.91

T3-T4 CORRELATIONS

T3AGE	N	PSI	PPVT	MOTOR	POCL1T	POCL1S
37-42	195-216	.57	.67	.37	.21	.28
43-48	227-244	.71	.78	.51	.22	.30
49-54	265-291	.68	.77	.46	.15	.33
55-60	227-250	.77	.78	.51	.29	.42
37-60	914-999	.77	.80	.62	.25	.33

---

[1] Number of items in each instrument:

PSI:	32
PPVT:	90 (Although termination criterion may be reached prior to 90th item).
MOTOR:	12
POCL1T:	5
POCLIS:	4

TABLE 4

INTRA-TIME POINT CORRELATIONS AMONG PHASE III TESTS[1]

T3

	PSI	PPVT	MOTOR	POCL1T	POCL1S	POCL2T	POCL2S	AGE
PSI		.70	.56	.29	.21	.19	.19	.53
PPVT	.74		.42	.22	.17	.14	.20	.46
MOTOR	.59	.52		.29	.07	.29	.09	.52
POCL1T	.35	.26	.33		.35	.63	.26	.23
POCL1S	.21	.18	.08	.41		.10	.70	.07
POCL2T	.22	.18	.33	.68	.22		.18	.17
POCL2S	.18	.18	.12	.35	.73	.33		.05
AGE	.56	.51	.61	.23	.07	.20	.08	

T4

---

[1] $p < .01$ for all correlations. N = 1246 - 1340 for T3; 974 - 1045 for T4.

TABLE 5

INTRA-TIME POINT PARTIAL CORRELATIONS, CONTROLLING FOR AGE[1]

		PSI	*PPVT*	*MOTOR*	*POCL1T*	*POCL1S*	*POCL2T*	*POCL2S*	
	PSI		*.61*	*.39*	*.21*	*.21*	*.13*	*.18*	
	PPVT	.64		*.24*	*.13*	*.16*	*.07*	*.20*	
	MOTOR	.38	.30		*.20*	*.05*	*.24*	*.08*	
T3	POCK1T	.27	.17	.25		*.35*	*.61*	*.25*	T4
	POCL1S	.20	.17	.05	.41		*.09*	*.70*	
	POCL2T	.14	.10	.27	.67	.21		*.18*	
	POCL2S	.17	.16	.09	.34	.73	.33		

[1] $\underline{p}$ < .05 for all correlations. N = 1246 - 1340 for T3; 974 - 1045 for T4.

CHAPTER TWO: PRESCHOOL INVENTORY

The PSI has demonstrated excellent psychometric characteristics throughout its use in the NDCS (as well as in other national studies--see Bache 1975). In Phase II the complete 64-item version was used, but analysis indicated that the 32-item version (produced by SRI for Year 3 of the Head Start Planned Variation Study) could be used with little loss in precision and sufficient time savings to permit addition of the PPVT to the Phase III battery. Performance characteristics of the PSI have been quite uniform throughout the study, as Table 6 shows. Phase II and Phase III means, standard deviations, and reliability coefficients are very similar within age groups at corresponding time points.

One of the principal strengths of the PSI is its developmental sensitivity. Its correlation with age has ranged from .53 to .56 at different times and the variance in scores has always been good for a 32-item test. The means in Table 6 provide another perspective for appreciating this property of the PSI. For any given test point, adjacent 6-month age groups are consistently about three points apart--an average gain of 1/2 point per month.

The PSI's sensitivity to individual differences in general and age in particular would not be possible, of course, were it not for the test's excellent reliability. T3 and T4 alpha coefficients, reported in Table 3, were .84 and .83 respectively. The lower reliabilities for the age subsamples should not mislead the reader. With any developmentally sensitive test, grouping on the basis of age substantially reduces the variance in true scores in each group.

Table 6

COMPARISON OF PHASE II AND PHASE III PSI SCORES
(32-Item Version)

	PHASE II			PHASE III			
	Fall (T1)			Fall (T3)			N
Age	Mean	SD	Alpha	Mean	SD	Alpha	PII, PIII
37-42	9.8	4.2	.69	9.8	4.1	.71	208, 288
43-48	12.5	5.4	.81	12.6	5.1	.78	308, 337
49-54	16.3	5.9	.83	15.5	5.2	.78	366, 377
55-60	19.3	6.0	.84	18.8	5.3	.80	253, 315
	Spring (T2)			Spring (T4)			N[a]
Age	Mean	SD	Alpha	Mean	SD	Alpha	PII, PIII
43-48	14.4	4.9	.73	14.4	4.6	.71	215, 208
49-54	17.7	5.3	.79	17.0	5.4	.80	310, 242
55-60	21.0	5.2	.81	19.8	5.1	.79	347, 307
61-66[b]	22.4	5.2	.83	22.4	4.7	.78	157, 285

---

[a] At T2, all children within the age limits were tested. At T4, only those age-acceptable children were tested who were also tested at T3.

[b] At T2, the oldest-age subsample was only 61-63 months. Only 11 children were older than 63 months.

Lower within-group reliability coefficients are virtually assured. Coefficients for the total sample are the appropriate estimates for the NDCS because scores of children across the entire age span were aggregated to the center level for the effects analyses. That is, analyses were not conducted within age groups. Note, nonetheless, that the reliability coefficient is somewhat lower for the youngest age group at each testing point.

The PSI's stability, as reflected in the T3-T4 test-retest correlations, was also excellent. In fact, these correlations were almost too high. As Stanley (1971) and others have pointed out, the reliability of a raw difference score approaches zero as the test-retest correlation approaches the internal consistency coefficient for the instrument. The reason is that, under these conditions, less and less of the variance in the difference scores can be attributed to individual differences in true-score change; more and more of this variance must be attributed to the test's error in measuring an individual's true score to begin with. Although residual gain scores corrected for attenuation were used in the NDCS (rather than raw difference scores), this relationship still holds.* In fact the formula for estimating the reliability of such corrected residual scores is:

$$\frac{\alpha^2 - r_{12}^2}{1 - r_{12}^2}$$

---

*See Appendix A for a more thorough discussion of the reliability of various measures of change.

where "alpha" is the internal consistency of the test (assumed homogeneous across time points, as is the variance in scores), and r_{12} is the test-retest correlation. Applying this formula to the PSI coefficients for the total sample, a reliability estimate of .43 for child-level scores was obtained. Aggregation to the center level improves this: Goodrich and Singer (1979) reported a generalizability coefficient of .63 for center-level means.

One of the advantages gained by using the shorter version of the PSI was a substantially reduced ceiling effect--the 32 items in the final version are generally more difficult than the 32 which were dropped. Nonetheless, ceiling effects were not altogether eliminated as a potential concern, as illustrated in Figure 1. Although only seven of 285 children aged 61-66 months scored 31 or higher at T4, the upper end of the score distribution is clearly more dense than the lower end. More importantly, most children in this group are not evenly distributed in terms of age; only 22 percent are 65 or 66 months old. There are two important consequences of this fact. First, if there had been more children in the 65-66 month interval, the PSI ceiling effect would almost certainly have been more pronounced. Second, even the slight ceiling effect shown in Figure 1 could have been serious if the children tested in a few of the centers had happened to be the oldest children in the sample. Such, however, was not the case.

As mentioned earlier, cognitive performance was not assessed in terms of single-time-point test scores, nor by raw difference scores or even simple residual gain scores. Rather, residual gains corrected for attenuation due to unreliability were computed, then aggregated to the center level for analysis. A complete discussion of the construction of these gain scores and their characteristics

Figure 1

DISTRIBUTION OF T4 PSI SCORES FOR CHILDREN 61-66 MONTHS OLD

```
T4 SCORE ON 32-ITEM PRESCHOOL INVENTORY

  8. ***
  9. ****
 10. ***
 12. ***
 13. ****
 14. ********
 15. ******
 16. *************
 17. ******
 18. **************
 19. ***************
 20. ***********
 21. *********************************
 22. **********************
 23. ****************************
 24. ********************
 25. **************************
 26. ****************************
 27. ************************
 28. ****************
 29. ********
 30. *****
 31. ******
 32. ***
     I.........I.........I.........I.........I.........I
     0        10        20        30        40        50
     FREQUENCY

MEAN           22.4      STD DEV       4.7       MEDIAN          22.9
VALID CASES      285     MINIMUM-MAXIMUM POSSIBLE SCORE:  0-32
```

is presented by Goodrich and Singer (1979). In light of the foregoing discussion, it should be pointed out here that the resultant measure was uncorrelated with the mean age of the children tested in each center.

CHAPTER THREE: THE PEABODY PICTURE VOCABULARY TEST

The PPVT was added to the Phase III battery when it was learned during Phase II that factoring the 64-item PSI did not yield a specific subscore for language ability (or any other subscores, for that matter). In order to avoid a potential cultural bias thought to be a problem with the original 1959 version of the PPVT, SRI used a set of revised plates developed by the Educational Testing Service for the Head Start Longitudinal Study (Meissner, Shipman, and Gilbert, 1972). Furthermore (as reported in Section 1.1), only items 11-100 of the 150-plate set were administered.

The decision to eliminate the first ten items proved to be ill-advised, as Figure 2 demonstrates. There is a pronounced floor effect in the distribution of T3 PPVT scores for the youngest age groups which would certainly have been reduced, if not eliminated, had these items been included.

But for this exception, the psychometric quality of the PPVT data was excellent. There was no ceiling effect whatever* and the test's reliability (.96 at T3; .95 at T4) was even higher than expected.[1] Reliability, in fact, was so high, that even the strong .80 test-retest correlation would not substantially reduce the reliability

---

*What may at first appear to be some sort of ceiling effect in the Table 2 subsample means is in fact a result of an uneven age distribution. The differences in adjacent T3 means, for example, are 6.6, 6.0, and 5.4 starting with the youngest group. There were, however, only 48 children aged 59-60 months in the 55-60 month age group compared to 146 children aged 55-56. When means were computed for each two-month interval, then averaged again to six-month intervals, the corresponding differences in adjacent means were 6.5, 6.3, and 6.5.

Figure 2

DISTRIBUTION OF T3 PPVT SCORES FOR CHILDREN 37-42 MONTHS OLD

```
T3 PEABODY PICTURE VOCAB TEST

 1.  **************
 2.  ****************************
 3.  ********************
 4.  **********
 5.  ***************
 6.  ***************
 7.  **************
 8.  ************
 9.  ********
10.  **************
11.  ********
12.  *****
13.  **********
14.  *************
15.  **************
16.  **************
17.  ***********
18.  *******
19.  *****
20.  ********
21.  ***********
22.  ********
23.  *****
24.  *******
25.  ****
26.  ****
27.  **
28.  *****
29.  ***
30.  ***
31.  ***
32.  ***
33.  ***
35.  ***
36.  ****
37.  ***
39.  ****
40.  **
43.  **
     I.........I.........I.........I.........I.........I
     0         10        20        30        40        50
     FREQUENCY

MEAN            12.8        STD DEV        9.6      MEDIAN          10.9
VALID CASES      285        MINIMUM-MAXIMUM POSSIBLE SCORE:   1-90
```

of the corrected residual change scores, which was estimated to be .86 at the child level. The generalizability coefficient for center-aggregated scores was reported by Goodrich and Singer (1978) to be .58.

When the PPVT was added to the NDCS test battery, there was some uncertainty whether it would correlate too highly with the PSI to justify using it as a distinct dependent measure. When age was partialled out, however, this correlation turned out to be .64 at T3 and .61 at T4, indicating that only about 40 percent of the age-independent variance in either test is predictable from the other. Although both instruments are certainly cognitive measures, it seems clear from these results that they do not measure identical sets of constructs.

CHAPTER FOUR: THE SRI TEST OF MOTOR SKILLS

The MOTOR test was actually administered as separate fine motor and gross motor scales, as explained in Section 1.1. Among children 55-60 months old, however, 34 percent obtained perfect T3 fine-motor scores and 7 percent perfect gross motor scores. Even when the items were considered a single test, 13 percent of these children scored 11 or 12 out of a possible 12.

This sort of ceiling effect suggests that change scores on the MOTOR test might only have been valid for children less than 55 months old at T4, or less than 49 months old at T3. Even this restricted use, however, would have encountered problems of reliability. Alpha coefficients for children 37-48 months old at T3 and 43-54 months old at T4 were only .59 and .50, respectively, while the test-retest correlation was .52. Under these conditions, the estimated reliability of corrected residual gain scores on the MOTOR test is only .10*. Such scores, therefore, would appear to be invalid.

---

*The estimated reliability is .10 whether internal consistency is assumed to be homogeneous at both time points (and .50 is averaged with .59) or whether it is allowed to be heterogeneous (see Note 1). The "correction" mentioned with respect to the residual change scores is, throughout this paper, an attenuation correction for test reliability, in which the test-retest correlation is divided by the alpha coefficient.

CHAPTER FIVE: THE PUPIL OBSERVATION CHECKLIST

As reported in Section 1.1, principal components analyses of the POCL in the National Home Start Evaluation and in Phases II and III of the NDCS have consistently revealed two factors: Task Orientation, consisting of 5 items, and sociability, consisting of 4. During Phase III, a five-point rating scale was used for the POCL, yielding possible score ranges of 5-25 for Task Orientation and 4-20 for Sociability. The testers showed such a strong positive bias, however, that the former scale was completely unusable and the validity of the latter was threatened.

This problem is illustrated by Figure 3, which presents the distribution of T4 Task Orientation and Sociability scores for the oldest group of children. The bias in ratings is dramatic, the modal score for the group being the maximum possible score. The bias, moreover, is not restricted to older children: 41 percent of the total sample of 1054 children tested at T4 received the maximum possible Task Orientation score.

As can be seen in Figure 3, the distribution of Sociability ratings is not as skewed. For the oldest age group (as well as for the total sample), only 11 percent of the scores were the maximum attainable. The principal challenge to the validity of the Sociability gain scores, however, concerns their reliability.

During Phase III, the PSI and Fine Motor Tests were administered on Day 1 and the PPVT and the Gross Motor tests on Day 2. The same tester administered all tests to any given

Figure 3

DISTRIBUTION OF T4 POCL DAY1 TASK ORIENTATION AND SOCIABILITY SCORES FOR CHILDREN 61-66 MONTHS OLD

```
        T4 DAY1 TASK ORIENTATION

         8. *
         9. *
        11. **
        15. **
        16. **
        17. **
        18. **
        19. **
        20. ******
        21. *******
        22. *****
        23. ******
        24. ************
        25. **************************************
            I.........I.........I.........I.........I.........I
            0        40        80       120       160       200
            FREQUENCY


MEAN          23.3          STD DEV       2.7          MEDIAN         24.5
VALID CASES    287          MINIMUM-MAXIMUM POSSIBLE SCORE: 5-25
```

```
        T4 DAY1 SOCIABILITY

         4. ***
         5. ****
         6. ******
         7. *************
         8. *************
         9. ********************
        10. *********************
        11. *******************
        12. *********************
        13. **********
        14. ******************
        15. ************
        16. *********************************************
        17. *******************
        18. *******************************
        19. ****************
        20. ********************************
            I.........I.........I.........I.........I.........I
            0        10        20        30        40        50
            FREQUENCY


MEAN          14.0          STD DEV       4.3          MEDIAN         15.0
VALID CASES    286          MINIMUM-MAXIMUM POSSIBLE SCORE: 4-20
```

child and completed the POCL ratings on each day. The correlation between Sociability ratings on Day 1 and those on Day 2 was .73 at T3 and .70 at T4. These coefficients, however, are grossly inflated estimates of the true rate-rerate reliability, which would include the rater as such as a source of error. Only at T2 were children tested by two different testers, and thus rated by different raters on the POCL. Only T2 data, therefore, provide a realistic estimate of POCL reliability.

At T2, one tester administered the Matching Familiar Figures Test (MFF) and the Motor Inhibition Test (MIT) on Day 1. A second tester administered the PSI and the Verbal Memory scale of the McCarthy Scales of Children's Abilities (MCCV) during Days 2 and 3. Although the Day 2-Day 3 correlation for Sociability was .71, the Day 1-Day 2 and Day 1-Day 3 correlations were both .44. This contrast clearly demonstrates the unreliability introduced when different raters are used to provide Sociability data, and indicates that rater variance is an appreciable component of the total variance in Sociability ratings.

Table 7 presents Sociability correlations within all four time points. It should be noted that all of the correlations involving the same rater on both days are within a range of .70-.73. This consistency is important because each correlation represents a pair of days on which slightly different batteries of tests were administered. At T1, for example, the .73 correlation reflects ratings made after the PSI1 was administered on Day 1 and the MFF and MCCV were given on Day 3. At T2, the .71 correlation reflects ratings made after the PSI2 was administered on Day 2 and the PSI1 and MCCV were given on Day 3. The consistency of these correlations despite differences in the

TABLE 7

TWO-DAY AND FALL-SPRING RATE-RERATE CORRELATIONS FOR POCL SOCIABILITY[1]

T1	Day 1: PSI1 Day 2: (No POCL) Day 3: MFF, MCCV	Same Tester/Rater

Correlation	Day 3
Day 1	.73

T2	Day 1: MFF, MOT	Tester/Rater 1
	Day 2: PSI2 Day 3: PSI1, MCCV	Tester/Rater 2

Correlations	Day 2	Day 3
Day 1	.44	.44
Day 2		.71

T3, T4	Day 1: PSI1, FMOTOR Day 2: PPVT, GMOTOR	Same Tester/Rater

T3 Correlation	Day 2
Day 1	.73

T4 Correlation	Day 2
Day 1	.70

T1-T2 Correlations

		T2		
		Day 1	Day 2	Day 3
T1	Day 1	.40	.42	.42
	Day 3	.39	.37	.42

T3-T4 Correlations

		T4	
		Day 1	Day 2
T3	Day 1	.33	.35
	Day 2	.35	.37

---

[1] Variable Abbreviations:

PSI1: 32-Item Preschool Inventory
PSI2: Remaining 32 items of the PSI
MFF: Matching Familiar Figures Test
MCCV: McCarthy Scales of Children's Abilities: Verbal Memory
MOT: Motor Inhibition Test

batteries implies that the low .44 correlations at T2 did not result from any peculiar interaction between the tests being administered and the ratings made afterward. Simply put, .44 is the best estimate obtainable from the NDCS data of the "true" reliability of Sociability ratings--including as many appropriate sources of error variance as can be included in one coefficient.

Table 7 also presents Sociability correlations between time points for Phase II and for Phase III. T1-T2 correlations range from .37 to .42, with a median of .41. T3-T4 correlation range from .33 to .37, with a median of .35. If the Phase II median of .41 is used in conjunction with the T2 reliability of .44, the estimated reliability of a corrected residual Sociability gain score turns out to be .09. Although an argument (admittedly not very strong) could be made for Sociability as a single-time-point measure, there appears to be no justification for computing a Sociability gain score.

CHAPTER SIX: CONCLUSIONS

On the basis of the information reported above and in other Volume IV papers, the following decisions were made with respect to the NDCS test battery.

- PSI and PPVT gain scores (corrected for attenuation) were computed and aggregated to the center level. Cognitive effects were evaluated with respect to these center-level averages.
- The SRI Test of Motor Skills was not used because the resulting gain scores would have been unreliable (and therefore invalid).
- The Pupil Observation Checklist was not used, because one of the scales was severely skewed and the reliability of gain scores computable for the other scale would have been suspect.

Other researchers using these instruments would be well-advised to note several specific concerns which affected these decisions. First, the reliability of change scores can and should be investigated empirically in any given research application. Second, the unit of analysis (level of aggregation) issue can also be examined empirically in a logical decision-making process. Third, it is important to focus attention on the distributional qualities of test data as well as on traditional properties such as reliability, mean performance and variability.

NOTE

[1] It should be noted in passing that the upper limit of the estimated internal consistency of a test is less than 1.0 when the item difficulties are heterogeneous. This is the case for the PPVT, whose first items are assumed to be passed by almost all children and whose last items are expected to be passed by none. Although the former assumption was unjustified (because of the elimination of items 1-10), the latter expectation was upheld. The following discussion provides some background for this issue, and concludes with specific computations for the PPVT based on Phase III data.

Stanley (1971) shows that the maximum value of KR20 when item difficulties are rectangular and item intercorrelations are maximized is

$$\text{Max KR20} = \frac{I}{I+1} \quad 1 - \frac{I(I+2)/6(I+1)}{I(I+2)/12} = \frac{I}{I-1} \quad \frac{I-1}{I+1} = \frac{I}{I+1} \qquad (1)$$

This is the general form of the KR20 formula, in which I is the number of items, I(I+2)/6(I+1) is the sum of the item variances and I(I+2)/12 is the variance in total scores under the conditions specified. It should be pointed out that KR20 is identical to coefficient alpha when items are scored dichotomously.

Now suppose that I=J+K, in which only K of the items have difficulties other than 0 and 1, so that the other J items have zero variances. In this case, the sum of the item variances is based only on the K items and is K(K+2)/6(K+1). Furthermore, any given total score is simply the score for the K items (X_k) plus a constant equal to the number of items whose level of difficulty (proportion passing) is 1.0. The variance in total scores (X), therefore is equal to the variance in the subscores X_k, and under the conditions specified is K(K+2)/12.

The maximum KR20 for a test of I items, therefore, when only K items have non-zero variances, and these variances are rectangularly distributed is:

$$\text{Max KR20} = \frac{I}{I+1} \quad 1 - \frac{K(K+2)/6(K+1)}{K(K+2)/12} \quad = \frac{I}{I-1} \quad \frac{K-1}{K+1} \qquad (2)$$

At both T3 and T4, 85 of the 90 PPVT items had non-zero variances. If the distribution of their variances had been perfectly rectangular, the upper limit of coefficient alpha at both time points (being the same as the maximum KR20) would have been:

$$\text{Max } \alpha = \frac{90}{89} \quad \frac{85-1}{85+1} = .988.$$

REFERENCES

W. Bache, A psychometric evaluation of the Preschool Inventory based on national research studies. Unpublished paper of the National Day Care Study. Cambridge, Mass.: Abt Associates Inc., October 7, 1975.

Educational Testing Service (Cooperative Tests and Services), Preschool Inventory Revised Edition--1970: Handbook. Princeton, N.J.: 1970.

R.L. Goodrich and J.D. Singer, Cognitive change in the National Day Care Study. In Vol. IV of the National Day Care Study Final Report. Cambridge, Mass.: Abt Associates Inc., December 1979.

J.A. Meissner, V.C. Shipman, and L.E. Gilbert, Technical Report 15: Peabody Picture Vocabulary Test, in V.C. Shipman (ed.), Disadvantaged children and their first school experiences: ETS Longitudinal Study Technical Report Series. Princeton, N.J.: Educational Testing Service, 1972. ETS Report No. PR-72-27.

V.C. Shipman, J.D. McKee, and B. Bridgeman, Disadvantaged Kids and their first school experiences: stability and change in family status, situational, and process variables and their relationship to children's cognitive performance. Report to the Office of Child Development, DHEW (Grant No. H-8256), September 1976. Princeton, N.J.: Educational Testing Service (ETS NO. PR-75-28).

J. Stanley, Reliability. In R.L. Thorndike (ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1971 (2nd ed.).

APPENDIX A: THE RELIABILITY OF THREE MEASURES OF CHANGE

INTRODUCTION

It is essential for the following discussion to distinguish between linked and independent scores--this distinction arising from the conditions under which the measurements are obtained. Generalizability theory (Cronbach, et al., 1972) provides a useful way of approaching this issue, as summarized nicely by Cronbach and Furby (1970, pp.69-70):

> There is a universe of possible conditions of observation of X. . . . There is a universe of observations of Y, and we shall assume that these may be made under the same set of conditions _i_ that are used for X observations. Then observation X_i and Y_i made under the same condition _i_ are said to be linked, and observations X_i and $Y_{i'}$, made under different [but parallel] conditions are said to be independent.*

Two conditions which are commonly repeated and thus produce linked measures are the test form and the tester or observer. Technically, however, repetition of any condition which might influence measurements (such as a given level of background distraction, the testing room, etc.) produces linked scores, although the "link" in some instances might be considered trivial. (Stanley [1971, p.364] offers a useful list of the possible sources of variance in test scores.) The relevance of this issue to a discussion of difference scores lies in the proper specification of the model and in the composition of the covariances which result.

---

*"Independent" in this usage does not, of course, mean uncorrelated, but merely "independently observed".

In classical test theory, an observed X is postulated as consisting of a true score component and a single uncorrelated error component: $X = T_x + e_x$. Furthermore, the classical approach also postulates that the error terms of two independent measures are uncorrelated, so that

$$\sigma(X,Y) = \sigma(T_x+e_x, T_y+e_y)$$

$$= \sigma(T_x,T_y) + \sigma(e_x,e_y)$$

$$= \sigma(T_x,T_y).$$

When X and Y are linked, however, the error covariance is _not_ necessarily zero, because the observation conditions (_i_) are the same for both measures.* It is necessary to specify two error terms, _e_ and _f_, which are uncorrelated with each other and with the true scores. That is, $X = T_x + e_x + f_x$. As in classical test theory, $\sigma(e_x,e_y)$ is always zero, but $\sigma(f_x,f_y)$ is zero only when X and Y are independent, as defined above. When X and Y are linked, $\sigma(f_x,f_y)$ 1) will usually not be zero; 2) may be large or small, depending on the extent to which the condition _i_ (common to both measurements) influences the X and Y scores; and 3) may be negative under certain circumstances (Cronbach and Furby, 1970, p.70). Formally, this model specifies that the set of conditions _i_ is sampled from a universe of parallel conditions. When a single _i_ is used to obtain both X and Y scores for a given person, the scores are said to be linked--the person's $\underline{f}_x$ and $\underline{f}_y$ score components are sampled simultaneously.

---

*More precisely, _i_ for the linked case refers only to those conditions which are the same for both X and Y. The conditions which are not identical need not be included in _i_ because errors from this subset are uncorrelated.

It is important to note the assumption's underlying the model. These are:

1) Errors are random and independent, except that when X and Y are linked, $\sigma(f_x,f_y)$ is usually not zero. With respect to any $\underline{e}_x$ component, its mean across persons is zero, its variance across persons is the same for every set of parallel conditions ($\underline{i}$, $\underline{i}'$, etc.), and its intercorrelation with T_x and f_x is zero. The same is true for any $\underline{e}_y$; and $\sigma(e_x,e_y) = \sigma(e_x,f_y) = \sigma(e_y,f_x) = 0$. Zero means, equal variances, and zero correlations with true scores are also assumed for $\underline{f}_x$ and $\underline{f}_y$; and $\sigma(f_x,f_{y'}) = 0$ when X and Y' are obtained under sets of conditions $\underline{i}$ and $\underline{i}'$ which are parallel but independently drawn from the universe.

2) Measures of X and X' made under different sets of parallel conditions $\underline{i}$ and $\underline{i}'$ are parallel. Therefore, from the previous assumptions, it follows that X and X' have equal means, equal variances, and equal intercorrelations with other measures. The same is held for Y and Y' measures.

3) It follows for all unlinked observations X and Y' that

$$\sigma(X,Y') = \sigma(T_x,T_y).$$

For linked observations, $\sigma(f_x,f_y)$ is assumed to be the same for all parallel sets of conditions. The covariance of linked X and Y measures, therefore, is the same under all parallel conditions:

$$\sigma(X,Y) = \sigma(T_x,T_y) + \sigma(f_x,f_y).$$

The notion of parallel conditions is a straightforward extension of the concept of parallel test forms, the test form being only one of the conditions of measurement. As with parallel forms, parallel conditions are those for which the true-score distributions overlap compeletely and the proportion of true-score variance to total variance is the same (Stanley, 1971, p.369). A set of conditions *i* consists not only of the variables which may affect a person's score (such as test form, tester, background noise level, etc.), but also the specific identity or "level" of each variable as it were. A parallel set of conditions must consist of the same variables which, on being sampled, possess the same properties with respect to their impact on measurement.

This is perhaps better understood by analogy to parallel test forms. In order to produce scores with equal variances, means, and correlations with other measures, the items in such forms must possess nearly identical psychometric properties, yet nonetheless be distinct from one form to the next. If the mean item difficulty or the variance in difficulties is different, the obtained scores will not have the properties required of parallel forms. Likewise, parallel sets of measurement conditions must exhibit identical psychometric properties in order to yield scores with the required characteristics. It would be no more permissable, for example, for background noise level to be 20 decibels in one condition and 90 decibels in another than it would be for the mean item difficulty to be .3 on one form and .6 on another.

There are, of course, flaws in the analogy between test items and measurement conditions. Psychometric properties of items, such as level of difficulty, for example, are equally relevant to all the items in a test.

They are well-known, and a large part of test theory is devoted to understanding their role in measurement. Measurement conditions such as tester (or observer) and testing room, however, are qualitatively distinct. Unlike items, the properties affecting obtained scores are different for each condition variable, are not as well known, and have not been as thoroughly investigated as have item characteristics. Consequently, the methodology required to produce parallel conditions of measurement is far less refined than that required to produce parallel forms.

A second flaw in the analogy is the disparity between the number of items in a form and the number of variables in a set of conditions. Tests usually consist of a relatively large number of items, so that slight departures from the requirement of identical properties can be tolerated as long as the requirement is met on the whole. In sampling item pairs from a large pool of pretested items whose properties are known, for example, it would not usually be considered necessary to achieve perfect matching in each pair in order to obtain reasonably parallel forms. The number of measurement conditions in a given experimental design, however, may be quite small, so that slight differences in the "level" of a single condition may have an important effect on the measurements obtained. If one actually wished to construct parallel sets of conditions for the purpose of obtaining unlinked measures, great care would need to be taken in specifying the precise details of the measurement methods. This would be particularly difficult to achieve in the case of ratings and naturalistic observations, which are known to be more susceptible than tests to systematic biases in the person providing or recording the data. (See, for example, Bache, 1979, and Connell, Goodson and Bache, 1977.)

The point to keep in mind with respect to the derivations which follow is that the different sets of conditions _i_ and _i_' referred to in the assumptions are not different in any way that affects the distributions of the observed X and X' scores, their reliability, or their correlation with other measures (such as Y and Y').

Notation

Cronbach and Furby's (1970) notation for distinguishing between the covariance of unlinked measures and the covariance of linked measures will be used.

$$\circ\,\sigma(X,Y) = \sigma(T_x,T_y) \tag{1}$$

will refer to the covariance of independently observed X and Y scores.

$$\bullet\,\sigma(X,Y) = \sigma(T_x,T_y) + \sigma(f_x,f_y) \tag{2}$$

will refer to the covariance of linked X and Y measures. Note that

$$\bullet\,\sigma(X,Y) = \circ\sigma(X,Y) + \sigma(f_x,f_y). \tag{3}$$

A similar distinction must be made between the correlation of independent measures ($\circ\rho_{xy}$) and the correlation of linked measures ($\bullet\rho_{xy}$).

$$\bullet\rho_{xy} = \frac{\circ\,\sigma(x,y)}{\sigma_x\ \sigma_y}, \tag{4}$$

whereas

$$\bullet \rho_{xy} = \frac{\bullet\sigma(X,Y)}{\sigma x \ \sigma y} = \frac{o\sigma(X,Y) + \sigma(f_x,f_y)}{\sigma x \ \sigma y}$$

$$= \frac{o\sigma(X,Y)}{\sigma_x \ \sigma_y} + \frac{\sigma(f_x,f_y)}{\sigma_x \ \sigma_y}$$

$$= o\rho xy + \frac{\sigma(f_x,f_y)}{\sigma_x \ \sigma_y} \qquad (5)$$

Since $\sigma(f_x,f_y)$ is usually, but not inevitably, positive, $\bullet\rho_{xy}$ is usually larger than $o\rho_{xy}$.

It is worth repeating, however, that the magnitude of $\sigma(f_x,f_y)$ depends on the influence the linked conditions of measurement have on the X and Y scores. In some cases, this influence may be trivially small, relative to the variance of X and Y, although a complex experimental design would be required to verify this empirically.

$\rho_{xx'}$ will represent the reliability of X*:

$$\rho_{xx'} = \frac{\sigma(X,X')}{\sigma_x^2} = \frac{\sigma_T^2}{\sigma_x^2} \qquad (6)$$

Reliability is usually defined formally in terms of parallel forms (X and X') whose true-score distributions and error variances are identical and whose errors are uncorrelated. Uncorrelated error implies not merely parallel forms but also an independent sampling from the universe of parallel conditions under which the measurements are made. In terms of the previous discussion, that is, X and X' are not linked.

---

*Some notational systems represent $\rho_{xx'}$ as $\rho^2_{T_x X}$
These are equivalent terms.

$\rho_{DD'}$ will represent the reliability of a difference score, however computed. Since D' is not linked to D, their covariance upon expansion will be seen to yield covariances between unlinked terms.*

All of the reliability derivations which follow assume that the X and Y measures from which D is computed are linked: $\bullet\rho_{DD'}$. Corresponding equations for the case of unlinked X and Y ($\circ\rho_{DD'}$) are also given, but without derivation. The three change measures which will be addressed are (1) raw difference scores, (2) simple residual scores, and (3) residual scores corrected for unreliability in the pretest.

The Reliability of Raw Difference Scores (Y-X) for Linked Variables

$$\bullet\rho_{DD'} = \frac{\sigma(D,D')}{\sigma^2_D} = \frac{\sigma[(Y-X),(Y-X)']}{\sigma^2_{(y-x)}}$$

$$= \frac{\sigma(X,X') + \sigma(Y,Y') - 2\sigma(X,Y')}{\sigma^2_x + \sigma^2_y - 2\bullet\sigma(X,Y)}$$

$$= \frac{\rho_{xx'}\sigma^2_x + \rho_{yy'}\sigma^2_y - 2\circ\rho_{xy}\sigma_x\sigma_y}{\sigma^2_x + \sigma^2_y - 2\bullet\rho_{xy}\sigma_x\sigma_y} \; . \qquad (7)$$

---

*D and D' are not linked even if the scores used to compute either one are linked. D and D', that is, are defined as being difference scores obtained independently under parallel conditions. The X and Y scores obtained in each instance, however, may be linked or unlinked depending on the experimental design.

If $\rho_{xx'} = \rho_{yy'}$ and $\sigma_x = \sigma_y$:

$$
{}^{\bullet}\rho_{DD'} = \frac{2\rho_{xx'}\sigma_x^2 - 2{}^{\circ}\rho_{xy}\sigma_x^2}{2\sigma_x^2 - 2{}^{\bullet}\rho_{xy}\sigma_x^2}
$$

$$
= \frac{\rho_{xx'} - {}^{\circ}\rho_{xy}}{1 - {}^{\bullet}\rho_{xy}} \; . \qquad (8)
$$

Substituting ${}^{\bullet}\rho_{xy} - \sigma(f_x, f_y) / \sigma_x^2$ for ${}^{\circ}\rho_{xy}$ in (8), we have

$$
{}^{\bullet}\rho_{DD'} = \frac{\rho_{xx'} - {}^{\bullet}\rho_{xy} + \sigma(f_x, f_y) / \sigma_x^2}{1 - {}^{\bullet}\rho_{xy}}
$$

$$
= \frac{\rho_{xx'} - {}^{\bullet}\rho_{xy}}{1 - {}^{\bullet}\rho_{xy}} + \frac{\sigma(f_x, f_y) / \sigma_x^2}{1 - {}^{\bullet}\rho_{xy}} \; . \qquad (9)
$$

The equation corresponding to (9) for the case of unlinked x and y measures is

$$
{}^{\circ}\rho_{DD'} = \frac{\rho_{xx'} - {}^{\circ}\rho_{xy}}{1 - {}^{\circ}\rho_{xy}} \; . \qquad (10)
$$

It may be noted from equation (10) that as the test-retest correlation between unlinked scores approaches the test's reliability, the reliability of raw differences between the unlinked scores approaches zero. It is also clear from equation (9) that the reliability of differences between linked scores would also approach zero as the linked test-retest correlation (${}^{\bullet}\rho_{xy}$) approached $\rho_{xx'}$, were it not for the correlated error which results from linking the measures. In fact, when ${}^{\bullet}\rho_{xy} = \rho_{xx'}$,

$$\rho_{DD'} = \frac{\sigma(f_x,f_y)/\sigma_x^2}{1 - \rho_{xx'}} \qquad (11)$$

It may seem strange that the reliability of a difference score should worsen as the test's stability improves. The reason, however, is that a greater and greater proportion of the variance in the difference scores is due to errors of measurement under these circumstances. This might better be understood by a short heuristic explanation. The explanation is stated in terms of unlinked scores, but the essential thrust of thea argument is the same for either case.

Consider a test with perfect reliability ($\rho_{xx'} = 1$). Observed scores obtained with such an insturment are, in fact, the true scores of those tested--there is no measurement error. If parallel administrations of alternate forms were conducted at two time points, and if each person either did not change or changed by the same amount as all the others, then the test-retest correlation would likewise be perfect ($r_{xy} = 1$). To the extent that the individuals did not all change by a constant amount, the test-retest correlation would be less than one. Thus, with a perfectly reliable test, the test-retest correlation would in fact be the index of individual differences in true-score change. The reliability of the observed difference scores would be perfect since the observed difference scores would actually be the true difference scores.

If a test is less than perfectly reliable, then observed scores reflect true scores plus error of measurement, and this error further contributes to diminishing the test-retest correlation. Since the reliability coefficient is an index of the error of measurement, while the unlinked test-retest coefficient is an index of true change plus error, then the difference between the two indicates the degree to which individual differences in observed change scores reflect individual differences in true change scores *over and above* the error component. In other words, the difference between test reliability and test-retest stability indicates the extent to which the variation in observed change scores is due less to error and more to genuine individual differences in true change.

As a test's stability approaches its reliability, less and less of the variance in observed change scores can be attributed to individual differences in true-score change. More and more of this variance must be attributed to the test's error in measuring an individual's true score to begin with. When $\rho_{xy} = \rho_{xx'}$, *all* of the variance in observed difference scores must be attributed to errors of measurement, and none can be assumed to reflect genuine differences in true scores. In short, the ability of the measure to detect real change has caught up with its ability to measure the variable it is supposed to measure, so that the reliability of the difference scores one could compute is zero. (When the difference score is computed from linked variables, only the fact that some of the *error* at the two time points is correlated prevents the linked, as well as the unlinked, reliability from actually reaching zero. The variance in *true* difference scores, nonetheless, is zero.)

The Reliability of Simple Residual Change Scores For Linked Variables

Residual change scores take the form

$$D = (Y - \overline{Y}) - \beta_{y.x}(X - \overline{X}),$$

or

$$D = (Y - \bullet\beta_{y.x}X) - (\bar{Y} - \bullet\beta_{y.x}\bar{X}), \tag{12}$$

in which $(\bar{Y} - \bullet\beta_{y.x}\bar{X})$ is a constant. The regression coefficient $\bullet\beta_{y.x}$ equals $\bullet\rho_{xy}\sigma_y / \sigma_x$. Letting the constant term in (12) be represented by *K*, the reliability of simple residual change scores for linked measures can be derived as follows.

$$\bullet\rho_{DD'} = \frac{\sigma(D,D')}{\sigma^2_D} = \frac{\sigma(Y - \bullet\beta_{y.x}X - K)(Y - \bullet\beta_{y.x}Y - K)'}{\sigma^2_{(Y - \bullet\beta_{y.x}X - K)}}$$

$$= \frac{\sigma(Y,Y') + \bullet\beta^2\sigma(X,X') - 2\bullet\beta\sigma(X',Y)}{\sigma^2_y + \bullet\beta^2\sigma^2_x - 2\bullet\beta\bullet\sigma(X,Y)}$$

$$= \frac{\rho_{yy'}\sigma^2_y + \bullet\beta^2\rho_{xx'}\sigma^2_x - 2\bullet\beta\circ\rho_{xy}\sigma_x\sigma_y}{\sigma^2_y + \bullet\beta^2\sigma^2_x - 2\bullet\beta\bullet\rho_{xy}\sigma_x\sigma_y}. \tag{13}$$

If $\rho_{xx'} = \rho_{yy'}$ and $\sigma_x = \sigma_y, \bullet\beta_{y.x} = \bullet\rho_{xy}$. Then:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'}\sigma^2_x + \bullet\rho^2_{xy}\rho_{xx'}\sigma^2_x - 2\bullet\rho_{xy}\circ\rho_{xy}\sigma^2_x}{\sigma^2_x + \bullet\rho^2_{xy}\sigma^2_x - 2\bullet\rho^2_{xy}\sigma^2_x}$$

$$= \frac{\rho_{xx'} + \bullet\rho^2_{xy}\rho_{xx'} - 2\bullet\rho_{xy}\circ\rho_{xy}}{1 - \bullet\rho^2_{xy}}. \tag{14}$$

Substituting $\bullet\rho_{xy} - \sigma(f_x,f_{y'})/\sigma_x^2$ for $\circ\rho_{xy}$, equation (14) can be extended as follows:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} + \bullet\rho_{xy}^2\rho_{xx'} - 2\bullet\rho_{xy}\,(\bullet\rho_{xy} - \sigma(f_x,f_y)\,/\sigma_x^2)}{1 - \circ\rho_{xy}^2}$$

$$= \frac{\rho_{xx'} + \bullet\rho_{xy}^2\rho_{xx'} - 2\bullet\rho_{xy}^2 + 2\bullet\rho_{xy}\sigma(f_x,f_y)\,/\sigma_x^2}{1 - \bullet\rho_{xy}^2}$$

$$= \frac{\rho_{xx'} - \bullet\rho_{xy}^2\,(2 - \rho_{xx'})}{1 - \bullet\rho_{xy}^2} + \frac{2\bullet\rho_{xy}\sigma(f_x,f_y)\,/\sigma_x^2}{1 - \bullet\rho_{xy}^2}\,. \qquad (15)$$

The equation corresponding to (15) for the case of unlinked X and Y measures is

$$\circ\rho_{DD'} = \frac{\rho_{xx'} - \circ\rho_{xy}^2\,(2 - \rho_{xx'})}{1 - \circ\rho_{xy}^2}\,. \qquad (16)$$

Only a slight extension of (15) and (16) is required to show that the reliability of a simple residual change score is not zero when the corresponding test-retest correlation ($\bullet\rho_{xy}$ or $\circ\rho_{xy}$) equals the test's reliability ($\rho_{xx'}$). For linked variables, $\bullet\rho_{xy} = \rho_{xx'}$ yields

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - 2\rho_{xx}^2 + \rho_{xx'}^3}{1 - \rho_{xx'}^2} + \frac{2\rho_{xx'}\sigma(f_x,f_y)\,/\sigma_x^2}{1 - \rho_{xx'}^2}$$

$$= \frac{\rho_{xx'}(1 - \rho_{xx'})^2}{1 - \rho_{xx'}^2} + \frac{2\rho_{xx'}\sigma(f_x,f_y)\,/\sigma_x^2}{1 - \rho_{xx'}^2}\,. \qquad (17)$$

As long as the test's reliability is not perfect, ${}_o\rho_{DD'}$ will be positive even if there is no correlated error. The same is true for unlinked variables, for which the extension of (16) yields

$$ {}_o\rho_{DD'} = \frac{\rho_{xx'}(1 - \rho_{xx'})^2}{1 - \rho^2_{xx'}} \, . \qquad (18)$$

If ${}_o\rho_{xy} = \rho_{xx'} = .5$, for example, then ${}_o\rho_{DD'} = .17$.
If ${}_o\rho_{xy} = \rho_{xx'} = .8$, ${}_o\rho_{DD'} = .09$.

It may seem counterintuitive that the reliability of a simple residual score remains positive under these conditions, especially when a lengthy explanation has just been offered as to why this shouldn't be the case. The reason for this curious result is that *observed* X and Y scores are used in computing the residuals, rather than *true* T_x and T_y scores. (See Note 14, p.390, Stanley, 1971). Note that the regression coefficient in (12) is $\beta_{y.x}$, not $\beta_{T_y.T_x}$. If the true-score coefficient *were* used, the reliability of unlinked residual scores would indeed by zero when ${}_o\rho_{xy} = \rho_{xx'}$.

This implies that correcting residual scores for test unreliability might produce a change score whose reliability *is* zero when the test-retest correlation equals the test's reliability. The final set of derivations shows this to be the case, but only for unlinked variables.

The Reliability of Residual Scores Corrected for Unreliability in the Pretest: Linked Variables

The test-retest correlation used in computing residual change scores is virtually always attentuated because of test unreliability. Corrected residual change scores address this problem by using $o\beta_{Y.X}/\rho_{xx'}$ instead of $o\beta_{Y.X}$ as the coefficient for predicting Y scores:

$$D = (Y - \frac{\bullet\beta_{y.x}}{\rho_{xx'}} X) - (\bar{Y} - \frac{\bullet\beta_{y.x}}{\rho_{xx'}} \bar{X}) \tag{19}$$

As in the case of residual scores, $\bullet\beta_{y.x} = \bullet\rho_{xy}\sigma_y / \sigma_x$. Letting K represent the constant $[\bar{Y} - (\bullet\beta_{y.x} / \rho_{xx'})\bar{X}]$ in (19), the reliability of corrected residual change scores for linked measures is derived as follows.

$$\bullet\rho_{DD'} = \frac{\sigma_{(D,D')}}{\sigma^2_D} = \frac{\sigma(Y - \frac{\bullet\beta_{y.x}}{\rho_{xx'}} X - K)\ (Y - \frac{\bullet\beta_{y.x}}{\rho_{xx'}} X - K)'}{\sigma^2\ (Y - \frac{\bullet\beta_{Y.X}}{\rho_{xx'}}\ X - K)}$$

$$= \frac{\sigma(Y,Y') + \frac{\bullet\beta^2}{\rho_{xx'}} \sigma(X,X') - 2\ \frac{\bullet\beta}{\rho_{xx'}}\ (X,Y')}{\sigma^2_Y + \left(\frac{\bullet\beta}{\rho_{xx'}}\right)^2 \sigma^2_x - 2\ \frac{\bullet\beta}{\rho_{xx'}}\ \bullet\sigma(X,Y)}$$

$$= \frac{\rho_{yy'}\sigma^2_y + \frac{\bullet\beta^2}{\rho^2_{xx'}}\rho_{xx'}\sigma^2_x - 2\ \frac{o\beta}{\rho_{xx'}}\ o\rho_{xy}\sigma_x\sigma_y}{\sigma^2_y + \frac{\bullet\beta^2}{\rho^2_{xx'}}\sigma^2_x - 2\ \frac{\bullet\beta^2}{\rho_{xx'}}\ \bullet\rho_{xy}\sigma_x\sigma_y} \tag{20}$$

If $\rho_{xx'} = \rho_{yy'}$ and $\sigma_x = \sigma_y$, $\bullet\beta_{y.x} = \bullet\rho_{xy}$. Then:

$$\circ\rho_{DD'} = \frac{\rho_{xx'}\sigma_x^2 + \frac{\bullet\rho_{xy}^2}{\rho_{xx'}}\sigma_x^2 - 2\frac{\bullet\rho_{xy}}{\rho_{xx'}}\circ\rho_{xy}\sigma_x^2}{\sigma_x^2 + \frac{\bullet\rho_{xy}^2}{\rho_{xx'}^2}\sigma_x^2 - 2\frac{\bullet\rho_{xy'}^2}{\rho_{xx'}}\sigma_x^2}$$

$$= \frac{\rho_{xx'} + \bullet\rho_{xy}^2/\rho_{xx'} - 2\bullet\rho_{xy}\circ\rho_{xy}/\rho_{xx'}}{1 + \bullet\rho_{xy}^2/\rho_{xx'}^2 - 2\bullet\rho_{xy}^2/\rho_{xx'}}$$

$$= \frac{\rho_{xx'} + \bullet\rho_{xy}^2/\rho_{xx'} - 2\bullet\rho_{xy}\circ\rho_{xy}/\rho_{xx'}}{1 - \bullet\rho_{xy}^2(2/\rho_{xx'} - 1/\rho_{xx'}^2)}. \qquad (21)$$

Substituting $\bullet\rho_{xy} - \sigma(f_x,f_y)/\sigma_x^2$ for $\circ\rho_{XY}$, equation (21) can be extended as follows:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} + \bullet\rho_{xy}^2/\rho_{xx'} - (2\bullet\rho_{xy}^2 - 2\bullet\rho_{xy}\sigma(f_x,f_y)/\sigma_x^2)/\rho_{xx'}}{1 - \bullet\rho_{xy}^2\,(2/\rho_{xx'} - 1/\rho_{xx'}^2)}$$

$$= \frac{\rho_{xx'} + \bullet\rho_{xy}^2/\rho_{xx'} - 2\bullet\rho_{xy}^2/\rho_{xx'} + 2\bullet\rho_{xy}\sigma(f_x,f_y)/\sigma_x^2\,\rho_{xx'}}{1 - \bullet\rho_{xy}^2\,(2/\rho_{xx'} - 1/\rho_{xx'}^2)}$$

$$= \frac{\rho_{xx'} - \bullet\rho_{xy}^2/\rho_{xx'}}{1 - \bullet\rho_{xy}^2(2/\rho_{xx'} - 1/\rho_{xx'}^2)} + \frac{2\bullet\rho_{xy}\sigma(f_x,f_y)/\sigma_x^2\rho_{xx'}}{1 - \bullet\rho_{xy}^2(2/\rho_{xx'} - 1/\rho_{xx'}^2)}. \qquad (22)$$

The equation corresponding to (22) for the case of unlinked X and Y measures is

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - \circ\rho^{2}_{xy}/\rho_{xx'}}{1 - \circ\rho^{2}_{xy}(2/\rho_{xx'} - 1/\rho^{2}_{xx'})} \; . \tag{23}$$

It is apparent in (23) that $\circ\rho_{DD'} = 0$ when $\circ\rho_{xy} = \rho_{xx'}$. It is equally apparent in (22) that, as long as $\sigma(f_x, f_y)$ is greater than zero (i.e., $\bullet\rho_{xy} > \circ\rho_{xy}$), $\circ\rho_{DD'}$ will not equal zero when $\bullet\rho_{xy} = \rho_{xx'}$. Under these conditions, in fact:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - \rho_{xx'}}{1 - \dfrac{2\rho^{2}_{xx'}}{\rho_{xx'}} - \dfrac{\rho^{2}_{xx'}}{\rho^{2}_{xx'}}} + \frac{2\,\sigma(f_x, f_y)/\sigma^{2}_{x}}{1 - \dfrac{2\rho^{2}_{xx'}}{\rho_{xx'}} - \dfrac{\rho^{2}_{xx'}}{\rho^{2}_{xx'}}}$$

$$= \frac{2\sigma(f_x, f_y)/\sigma^{2}_{x}}{2 - 2\rho_{xx'}}$$

$$= \frac{\sigma(f_x, f_y)/\rho^{2}_{X}}{1 - \rho_{xx'}} \; , \tag{24}$$

which is the same equation as that obtained for raw difference scores (11). As equation (24) shows, only the correlated error which results when variables are linked prevents $\circ\rho_{DD'}$ from being zero when the linked test-retest correlation equals the tests' reliability.

Summary

The foregoing derivations might best be summarized by presenting their end results for the cases of linked change scores.

For raw difference scores:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - \bullet\rho_{xy}}{1 - \bullet\rho_{xy}} + \frac{\sigma(f_x, f_y) \, / \, \sigma_x^2}{1 - \bullet\rho_{xy}} . \qquad (9)$$

For simple residual scores:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - \bullet\rho_{xy}^2 \, (2 - \rho_{xx'})}{1 - \bullet\rho_{xy}^2} + \frac{2\bullet\rho_{xy}\sigma(f_x, f_y) \, / \, \sigma_x^2}{1 - \bullet\rho_{xy}^2} . \qquad (15)$$

For residual scores corrected for unreliability in the pretest:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} - \bullet\rho_{xy}^2/\rho_{xx'}}{1 - \bullet\rho_{xy}^2(2/\rho_{xx'} - 1/\rho_{xx'}^2)} + \frac{2\bullet\rho_{xy}\sigma(f_x, f_y)/\sigma_x^2\rho_{xx'}}{1 - \bullet\rho_{xy}^2(2/\rho_{xx'} - 1/\rho_{xx'}^2)} . \qquad (22)$$

Finally, the corresponding reductions obtained when $\bullet\rho_{xy} = \rho_{xx'}$ are:

For raw difference scores:

$$\bullet\rho_{DD'} = \frac{\sigma(f_x, f_y)/\sigma_x^2}{1 - \rho_{xx'}} . \qquad (11)$$

For simple residual scores:

$$\bullet\rho_{DD'} = \frac{\rho_{xx'} \, (1 - \rho_{xx'})^2}{1 - \rho_{xx'}^2} + \frac{2\rho_{xx'}\sigma(f_x, f_y)/\sigma_x^2}{1 - \rho_{xx'}^2} . \qquad (17)$$

For residual scores corrected for unreliability in the pretest:

$$\bullet\rho_{DD'} = \frac{\sigma(f_y, f_x)/\sigma_x^2}{1 - \rho_{xx'}} . \qquad (24)$$

REFERENCES

Bache, W.L. III. The Dangers of Confounding Rater Bias with Group Differences, presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 1979.

Connell, D.C., Goodson, B.D., and Bache, W.L. III Observer/Tester Effects on Process and Dependent Measures, unpublished Research Memorandum of the National Day Care Study, Cambridge, Mass.: Abt Associates Inc., June 1, 1977.

Cronbach, L.J. & Furby, L. How We Should Measure "Change" - or Should We? Psychological Bulletin, 1970, 74, 68-80.

Cronbach, L.J., Gleser, G.C., Nanda, H., Rajaratnam, N., The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York: John Wiley and Sons, 1972.

Stanley, J., Reliability, in R.L. Thorndike (ed.), Educational Measurement, Washington, D.C.: American Council on Education, 1971 (2nd ed.).

Investigation of Teacher Rating Scales Considered for Use in the National Day Care Study

William L. Bache III

TABLE OF CONTENTS

Page

CHAPTER ONE: INTRODUCTION 105

  Method 106

CHAPTER TWO: WITHIN-PAIR t-TASKS AND CORRELATIONS 109

CHAPTER THREE: PERCENTAGE OF AGREEMENT BETWEEN RATERS 111

CHAPTER FOUR: CAREGIVER PAIR CHARACTERISTICS 114

CHAPTER FIVE: RECOMMENDATIONS 116

REFERENCES 117

CHAPTER ONE: INTRODUCTION

Behavioral ratings and checklists are an attractive measurement technique. They are simple to use, require little of the rater's time and none of the subject's, and can capitalize on the impressions of those most familiar with the subject. Virtually any list of behaviors, traits, or concepts can be adapted to a checklist or rating format which possesses a high degree of face validity.

On the other hand, these techniques are subject to serious flaws which are often not apparent during the design, administration, analysis, or interpretation of the scales. The behaviors to be assessed, for example, may suffer from a lack of specificity or objectivity and may be subject to different interpretations of what is to be included. Furthermore, even a list of behaviors which is specific must proceed from or be relatable to a theoretical framework which gives meaning to the data. Checklists are prone to be developed without sufficient regard for this fact; or worse yet, are prone to be developed with respect to a "theory" which consists solely of the author's opinions, prejudices and values regarding human behavior. In addition, ratings suffer from a number of commonly-known sources of error, including halo effects, rater leniency, ambiguity of trait labels or of scale point definitions, varying levels of information about the persons rated, varying rater response styles, errors of rater logic regarding trait relationships, and other sources of rater-trait and rater-subject interactions. It is these latter sources of rater error which are of special concern when different raters provide the data for different groups in a research design.

The purpose of the research reported here was to determine whether rating scales could be used to augment the test battery used in the National Day Care Study (NDCS). The design of the NDCS would have required the use of different raters in each of the study's day care centers, thus confounding all sources of rater error with genuine center differences. During June of 1976, the Stanford Research Institute (SRI) field tested four rating scales designed to allow caregivers to rate children on several social-emotional dimensions. SRI's preliminary report on this field test left a number of questions unanswered as to the usefulness of these instruments for Phase III of the NDCS. Specifically, it was not clear whether teachers whose ratings were highly correlated also agreed with each other in terms of the absolute magnitude of their scores. Secondly, there were no analyses indicating whether the characteristics of individual caregivers or caregiver pairs were related to interrater agreement on these instruments. This paper reports the following re-analyses of the data.

- An examination of *t*-tests and correlations for each subscale and rater pairs.
- An examination of the percentage agreement between raters.
- An examination of caregiver and caregiver-pair characteristics related to interrater agreement.

Method

Ten pairs of caregivers were selected in each of the three NDCS sites: Atlanta, Detroit, and Seattle. These thirty pairs represented thirty classrooms in 21 of the 64 day care centers which participated in the NDCS. Classrooms were selected to provide a mix of staff/child ratios, group sizes, and staff education representative of the entire NDCS sample. Caregivers were paired on the basis of the equivalence of their familiarity with the children to be rated. In all instances, the ten children which each pair rated had been enrolled in their classes for at least six months.

All caregivers participated in either half-day or full-day training sessions for one week, the half-day sessions being followed by a week of child observation before rating data were collected. The training was designed to acquaint the raters with common sources of error and bias, and attempted through practice and discussion to reduce these influences before the ratings made for this study were obtained. Subsequent analyses showed no advantage for either training method over the other and data from both have been pooled. The total data base consisted of information on 300 children: 30 pairs of raters each rating 10 children.

Four rating scales (yielding a total of eight subscales defined by the scale authors) were selected on the basis of their potential usefulness in augmenting the NDCS test battery. The items in all four scales consist of behavior descriptions which are rated in terms of their frequency of occurrence.

- Schaefer Day Care Behavior Inventory. A 30-item, five-point rating scale with three subscales: a) Introversion-Extroversion, b) Task Orientation-Distractability, and c) Considerateness. Items are rated "almost always" through "almost never."

- Kohn-Rosman Problem Checklist. A 49-item, three-point scale with two subscales: a) Apathy/Withdrawal and b) Anger/Defiance. Items are rated "not at all typical," "somewhat typical," and "very typical," with the latter two categories indicating the possible existence of problem behavior. This scale showed very low variance in item response, "not at all typical" being marked 80 to 90 percent of the time.

- Kohn-Rosman Social Competence Scale. A 73-item, five-point scale consisting of two subscales: a) Interest/Participation-Apathy/Withdrawal and b) Cooperation/Compliance-Anger/Defiance. Items are rated "very often or always" through "hardly ever or seldom."

- Vineland Social Maturity Scale. An 11-item scale which assesses self-help skills associated with toileting, eating, dressing, and avoiding hazards. Items are rated in terms of whether the child "usually or habitually" performs them with success, "occasionally but not usually" performs them, or "rarely or never" performs them. This scale, too, evidenced very restricted item variance. Almost all skills were performed "usually and habitually" by all the children rated.

CHAPTER TWO: WITHIN-PAIR t-TESTS AND CORRELATIONS

It is not sufficient, of course, to assess interrater agreement only by correlation because this compares raters only in relative terms. It is quite possible for differences in response style between two raters to yield absolute differences in their mean ratings for a group of children even when the ratings are perfectly correlated. For each pair of raters in the study, therefore, Table 1 presents both interrater correlations and the results of *t*-tests of mean differences between the paired raters. Given eight subscales and 30 pairs (each rating 10 children), it is hypothetically possible to compute 240 correlations and conduct 240 *t*-tests, but missing data and zero variances reduced this to 233. A difference between the means of two raters which is significant at the .05 level of probability is indicated by a "*t*" in Table 1 whereas correlations are reported at their actual computed value.

The bottom three lines in Table 1 summarize the analysis and indicate how problematical ratings would be for the NDCS. Depending on the subscale, fully one-half to two-thirds of the interrater correlations are below .70, and one-third to one-half of the rater differences are significant. When both criteria are considered jointly, only about one-quarter to one-third of the rater pairs demonstrate satisfactory correlation and no significant difference--and these are not the same rater pairs for each subscale. Overall, only 58 of the 233 pair-by-subscale analyses demonstrated acceptable interrater agreement. Clearly, these results do not support the use of these rating scales in the NDCS. Apart from the poor correlations, center differences in ratings would very likely be obtained strictly as an artifact of differences in response style.

Table 1

WITHIN-PAIR INTERRATER CORRELATIONS AND t-TESTS OF MEAN DIFFERENCES

RATER PAIR	BEHAVIOR INVENTORY						PROBLEM CHECKLIST				SOCIAL COMPETENCE				VINELAND	
	Extrov.		Task Or.		Consid.		Apathy		Anger		Int-par		Cooper			
	r	t	r	t	r	t	r	t	r	t	r	t	r	t	r	t
D 01	.70		.88		.88		.96		.38		.90		.80		.69	t
D 02	-.49		.41		.44		.32	t	.57		.38	t	.49		.45	t
D 03	.76		.90	t	.87	t	.28		.69		.78	t	.91		.80	t
D 04	-.31	t	.59	t	.74	t	.29	t	.06	t	.69	t	.45		.87	t
D 05	.28		.54		-.11		.48		.55		.70		.45		.83	
D 08	.33	t	.10		.33		-.03		.65		.28	t	.55		.87	t
D 09	.88	t	.65		.85		.71		.94		.94		.90	t	.77	
D 10	.91	t	.93		.38	t	.20		.73		.96	t	.85	t	.26	
D 11	.53		.82	t	.45		.67		.88		.82	t	.83		(Ovar.)	
D 12	.57	t	.69		.38		.85	t	.48	t	.86	t	.57		.81	
A 02	.96		.80		.88		.51	t	.62		.65		.83		.42	
A 03	.36		.53	t	.56	t	—		—		.77	t	.63	t	.65	
A 04	.84	t	.77		.74	t	.51		.67		.76	t	.46	t	.07	t
A 06	.15	t	.72	t	.52		—		—		.57	t	.50	t	.47	t
A 07	.66		.83		.75	t	.85	t	.56	t	.71		.76	t	.97	t
A 08	.62		.20		.56		-.03		.44		.34		.67		-.11	
A 09	.65		.62		.66		.40		.64	t	.81		.51	t	-.17	t
A 10	.82		.64		.99	t	.41		.96	t	.62		.97	t	.25	t
A 11	.75	t	.57		.64		.73		.81	t	.48	t	.65		(Ovar.)	
A 13	.37		-.03	t	.67		.06		.35		.14	t	.66	t	(Ovar.)	
S 03	.56		.65	t	.78		.28		.94		.69		.83	t	.60	t
S 04	.45		.64		.80		.76		.99	t	.55		.94		-.13	
S 05	.54		.58	t	.92		.64		.94	t	.56		.94		.85	t
S 06	.33		.89		.72		.48	t	.89		.89		.87	t	.06	
S 07	.26		.82		.78	t	.13		.58	t	.70		.77	t	.59	
S 08	.73		.71		.37		.64	t	.74		.86	t	.90		.63	
S 09	-.21	t	.81	t	.93		.95		.64	t	.60	t	.80	t	-.02	
S 12	.65	t	.77		.76		.72		.87		.83	t	.87		.67	
S 13	.76	t	.70	t	.66	t	.94	t	.64	t	.85	t	.70	t	.62	
S 14	.29		.67		.77		.81		.77		.50		.88		.71	t
OF 30 PAIRS: Number with r ≥ .70	10		14		16		10		12		16		18		9	
Number with non significant t-test	19		20		21		20		17		14		16		14	
Number with r ≥ .70 and non significant t-test	5		9		10		7		8		7		9		3	

Note: t indicates a mean difference between raters significant at the .05 level of probability.

CHAPTER THREE: PERCENTAGE OF AGREEMENT BETWEEN RATERS

The traditional method of computing the percentage of agreement between two raters is to divide the number of items on which the raters agree by the total number of items. When, however, item variance is low due to the rare occurrence of atypical behavior, the traditional index is spuriously inflated by agreement as to mere nonoccurrence, and this may mask interrater disagreement as to occurrence. A more appropriate measure in these circumstances is the percentage of agreement on amodal or atypical responses. This calculation is based on the number of items on which at least one of the raters uses one of the less used rating categories. Percentage of agreement is computed by dividing ths number into the number of items on which the other rater also indicates atypical behavior by the child. Essentially, this procedure reduces the rating scale to a yes-no checklist and determines the extent to which the raters agree on "yes" responses, while ignoring items on which both raters check "no." In one respect, it maximizes the "true" agreement index because it assesses only whether the raters agree as to the occurrence of atypical behavior, and not whether they concur as to the degree of atypicality.

Table 2 presents the results of both the traditional and the amodal percent agreement analyses. Behavior Inventory and Social Competence scales were rescaled to three points after the first round of analyses because the raters agreed perfectly only about 40 percent of the time with the five-point scaling, but were within one point of each other another 40 percent of the time. As it turned out, however, rescaling did not increase agreement as much as hoped, the traditional percent agreement increased only to about 60 percent with the three-point scaling. Furthermore, reducing the scale to three points often lowered the item variance.

Table 2

TRADITIONAL AND AMODAL PERCENTAGE OF AGREEMENT BETWEEN RATERS (AVERAGED OVER THIRTY PAIRS OF RATERS)

	TRADITIONAL: 5-point scaling		TRADITIONAL: 3-point scaling	AMODAL: 3-point scaling
	Percentage of Items 0 pts. apart	Percentage of Items 1 pt. apart	Percentage of Items 0 pts. apart	Percentage of Agreement on Any Amodal Response[c]
Behavior Inventory	41	36	63	38
Social Competence	42	38	62	38
Problem Checklist	a	a	79[b]	26
Vineland Social Maturity	a	a	84[b]	20

---

[a] The Problem Checklist and the Vineland are three-point scales to begin with.

[b] These percentages are spuriously inflated because of low variance in the item responses: on 80 to 90 percent of the items, every child was given the same rating.

[c] The amodal agreement percentage is obtained by dividing the total number of items on which both raters used one of the two amodal responses by the number of items on which either rater used an amodal response.

At first glance, it appears there is excellent interrater agreement on the Problem Checklist and the Vineland scales, the traditional percentage of agreement being about 80 percent for both. Closer inspection, however, revealed that this resulted primarily from low item variances in the scales. The percentage of agreement on amodal responses reveals how low the interrater agreement on all four scales really is. On no scale did raters agree more than 40 percent of the time when one or the other of them checked one of the two less frequently used rating categories. Such a low incidence of agreement does not speak well for the use of these scales in the NDCS.

CHAPTER FOUR: CAREGIVER PAIR CHARACTERISTICS

The 30 pairs of raters in the study can be sorted into a factorial design based upon the two dimensions of interrater agreement: correlations and absolute score agreement. Using the results in Table 1, the mean correlation for each caregiver pair and the number of nonsignificant t-tests over the eight subscales were used to classify rater pairs as "high" or "low" along the two dimensions. The 2x2 factorial design for this analysis is illustrated in Figure 1. Only 20 rater pairs (five in each cell) were included in the analysis in order that a balanced design might be maintained and that clear distinctions between cells be achieved. Ten caregiver and caregiver-pair characteristics (listed in Figure 2) were then subjected to two-way analysis of variance in order to determine whether rater agreement was related to such factors. Of the 30 possible significant F-tests, only one (an interaction for the size of the class) was significant, even at the .10 level of probability. There is, therefore, no data indicating that rater agreement is related to the rater and classroom characteristics listed in Figure 2.

Figure 1

Design for Analysis of Caregiver-Pair Characteristics

		Correlational Agreement	
		High	Low
Absolute Score Agreement (between means)	High	$\bar{r}$ = .76 ns = 6.4	$\bar{r}$ = .42 ns = 6.7
	Low	r = .75 ns = 2.9	r = .50 ns = 2.9

N = 5 rater pairs per cell

"$\bar{r}$" is the average interrater correlation for the five pairs of raters in the cell.

"ns" is the average number of t-tests (out of 8 possible) on which no significant differences between rater means were found. The higher the number the greater the agreement between raters in a pair.

Figure 2

Dependent Variables in Caregiver Pair Characteristics Analysis

1. Years of experience of Rater 1
2. Years of experience of Rater 2
3. Total years of experience of the pair
4. Difference in years of experience betwen raters in the pair
5. Years of education of Rater 1
6. Years of education of Rater 2
7. Total years of education of the pair
8. Difference in years of education between raters in the pair
9. Staff/Child ratio of the caregivers' class
10. Size of the caregivers' class

CHAPTER FIVE: RECOMMENDATIONS

These results do not support the use of rating scales in the NDCS--at least not as dependent measures of children. Neither interrater reliability (in terms of the correlation between raters) nor absolute interrater agreement (in terms of mean differences and percentage agreement) are adequate.

It is occasionally suggested that rating scales might be used as a source of information about the caregiver-rater rather than about the children. This would only be possible, however, if it were known that the children did not actually differ (at the level of aggregation used for analysis) along the dimensions being rated. Since this cannot be assumed, it is difficult to see how ratings could be used to assess the raters in the NDCS.

REFERENCES

Cronbach, L. J. Essentials of Psychological Testing. New York: McGraw-Hill, 1960, 2nd ed.).

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.

Guildford, J. P. Personality. New York: McGraw-Hill, 1959.

Nunally, J. C. Introduction ot Psychological Measurement. New York: McGraw-Hill, 1970.

An Analysis of the CDA Checklist Data

Nancy Goodrich

TABLE OF CONTENTS

Page

CHAPTER ONE: INTRODUCTION 121

CHAPTER TWO: DATA DESCRIPTION 123

- The CDA Checklist 123
- Data Reduction and Analytic Issues 124
- CDA Reliability 127

CHAPTER THREE: THE CDA ANALYSES 131

- Main Effects Analyses 131
- Classroom Process Analyses 136
- Child Test Score Analyses 139

CHAPTER FOUR: CONCLUSIONS AND RECOMMENDATIONS 142

REFERENCES 144

APPENDIX A 145

APPENDIX B 167

APPENDIX C 171

APPENDIX D 181

CHAPTER ONE: INTRODUCTION

Since children in day care spend much of their time with individual caregivers, it is important to investigate the relationships between caregiver skills and behaviors and children's growth and development. In the National Day Care Study (NDCS), caregiver behavior was recorded using two observation systems--the Adult-Focus Instrument (AFI) and the Child Development Associates (CDA) Checklist. Most of the NDCS analyses have focused on caregiver behavior as measured by the AFI (Goodson, 1978). The AFI information, however, was supplemented by the CDA Checklist, which was recorded immediately following the Adult Focus observation. Developed by Stanford Research Institute, the CDA Checklist identifies caregiver skills and behaviors relevant to eleven functional areas of caregiver competency defined in the Child Development Associates credentialing system.

This report is primarily concerned with the psychometric properties of data from the CDA Checklist. The results of some preliminary analyses that investigated relationships between policy variables, selected CDA variables and children's behaviors and cognitive gain scores also are presented. Major findings are as follows:

- For analysis purposes, CDA variables can be organized into four factors: ENVIRONMENT, RESOURCES, CLASS MANAGEMENT and CHILD ORIENTATION.

- Classroom structure (NUMBER OF CAREGIVERS, STAFF/CHILD RATIO) is associated with day care centers that are safe, sanitary and have many resources available and accessible to children.

- Caregiver qualifications (YEARS OF EDUCATION, SPECIALIZATION) are related to caregiver activities that are oriented toward child development and toward maintaining a well-organized classroom.

- Caregivers who direct their attention to children (CHILD ORIENTATION), are associated with children who are more involved and who show greater gains in child development, especially as measured by the PPVT.

The analyses behind these findings are summarized below. A copy of the checklist and other supporting materials are provided in Appendices A through D.

CHAPTER TWO: DATA DESCRIPTION

The CDA Checklist

The Child Development Associates credentialing system was developed by the CDA Consortium to certify caregivers who are competent in helping children learn and develop. To be accredited, candidates generally participate in an extensive period of CDA training. In addition, they must compile a portfolio containing samples of their work with children, be observed in the classroom working with children and be recommended by children's parents. Candidates must be at least 16 years old and have worked in a group setting with young children for at least eight consecutive months full-time (or 16 months part-time). A formal educational degree, however, is not required.

The CDA Consortium has defined six general areas in which a person should be proficient in order to work effectively with young children. For assessment purposes, these competency areas are further divided into thirteen functional areas that provide a framework for evaluating caregivers: safety, health, classroom environment, physical competence, cognitive development, language development, creative expression, self-concept, individual strengths, prosocial behavior, group management, home-center relationship and staff skills. The CDA Consortium, however, stresses the need to perform an individual and open-ended assessment of each candidate and repeatedly emphasizes that there are so many potential indicators of quality caregiver behavior that it is not possible to cite critical behavior that must be evidenced.

The open-ended assessments prescribed by the Consortium are an extremely useful basis for awarding a CDA credential to an individual caregiver. However, because a fair assessment may require more than a year to complete and may be difficult to quantify, a more structured approach to a CDA-based evaluation of caregiver behaviors and skills is required for analysis purposes. The CDA Checklist was developed primarily to investigate statistically the organization and definition of the CDA competency and functional areas. The information also was used in the NDCS to examine relationships between caregiver behaviors, based on the CDA definitions, and children's behaviors and cognitive gains.

The CDA Checklist consists of 235 items organized around eleven of the functional areas defined by the CDA Consortium;* many of the items included on the checklist were obtained from open-ended observations completed as part of the actual credentialing process. In addition, judgmental summary variables were developed for each area in an attempt to maintain the subjective aspect of the CDA system. The checklist, with instructions to observers, appears in Appendix A.

Data from a total of 261 checklists were included in the NDCS analyses. Caregivers were observed in spring 1977 on two different days by two different observers; the average response for the two days was computed to produce a single CDA profile for these caregivers. For those caregivers observed only once, the single response was used.

Data Reduction and Analytic Issues

Factor analyses were used to investigate the psychometric properties of the CDA data and to reduce the many checklist items to a more manageable data set. Because the CDA Checklist contains too many items to be included in a single factor analysis using currently available software, the initial analyses were directed toward - obtaining one or more significant factors in each of the eleven functional areas. Both principal components analysis and Rao's maximum likelihood factor analysis were performed using three samples: data from observations on day one, data from observations on day two, and checklist responses averaged across the two days. Similar factor loadings were found for all three samples. Based on these results, twelve factors were defined for further analysis (see Table 2.1); items included in each factor are shown in Appendix B.

---

*Home-center and staff areas could not be assessed during an observation period and therefore were omitted. In addition, cognitive and language development items were combined in the checklist organization, and classroom environment was included with safety and health.

Initial analysis showed high correlations among CDA factors. This finding is not unexpected. Even though the CDA Consortium defines thirteen functional areas, these areas are intended primarily to organize the assessments rather than to serve as independent measures of caregiver competence. To receive accredidation, a caregiver must perform well in all areas.

Because of the multicollinearity, additional factor analyses were performed on the twelve factors to investigate the existence of second-order factors. It was anticipated that such factors might correspond to the six Competency Areas defined by the CDA Consortium. The results (Table 2.1) indicated that three second-order factors--RESOURCES, CHILD ORIENTATION and CLASS MANAGEMENT--could be identified. A fourth factor, ENVIRONMENT, subsequently was defined to combine the two remaining factors, maintains "safe classroom" and maintains "sanitary classroom".

The four second-order factors reflect different but related aspects of caregiver skills and behaviors. RESOURCES describes the availability of materials and special areas in the classroom--tricycles, balls, doll houses, musical instruments, reading areas, science areas

Table 2.1

Factor Analysis of CDA Factors

Factors	RAO Loadings		
	1 (Resources)	2 (Child Orientation)	3 (Management)
Maintains Safe Classroom	.25	.10	.15
Encourages Safety	.06	.27	→ .79
Maintains Sanitary Classroom	.31	.21	.06
Provides Gross Motor Toys	→ .49	-.02	.09
Encourages Active Play	.06	→ .36	.06
Encourages Cognitive/Language Dev.	.14	→ .81	.26
Provides Creative Play Materials	→ .49	.18	.26
Encourages Good Self-Concept	.29	→ .64	.42
Encourages Self-Help	.29	→ .38	.13
Encourages Social Behavior	.20	→ .60	.43
Manages Class Activities Well	.44	.37	→ .66
Arranges Classroom Well	→ .87	.22	.03

and quiet areas. CHILD ORIENTATION reflects the caregiver's emphasis on child development and her interactiveness with children. Caregivers who scored high on this construct encouraged children to learn, play and interact well with others. The CLASS MANAGEMENT factor reflects how well the caregiver organizes and manages classroom activities--for example, safeguarding children, starting a new activity and restraining anger. The final factor, ENVIRONMENT, addresses the safety and cleanliness of the classroom. However, because many of the items reflected in this factor are likely to be the responsibility of the center director rather than of the classroom staff, ENVIRONMENT is probably best viewed as a center descriptor rather than as a measure of caregiver skills. In addition, the variance for this factor was not high, reducing its analytic usefulness.

The four second-order factors helped to simplify analysis and interpretation; all except ENVIRONMENT, however, remain significantly correlated with one another. Therefore an overall CDA rating (CDA SCORE) was constructed by taking an average of factors, weighted by the number of items in each factor. This rating may be viewed as a single measure of caregiver skills and behaviors that may be related to caregiver competence in the same way that obtaining a CDA credential reflects overall competence.

CDA Data Reliability

To investigate the reliability of the twelve first-order factors, classrooms were selected that had been observed by a black observer on one day and by a white observer on the other day (see Singer, Affholter and Goodrich, 1978). Factor scores were computed for each day, and correlations were calculated across days. Table 2.2 presents these correlations both by site and using a split (even/odd) sample. The results provide clear evidence that

Table 2.2

Day A - Day B Factor Correlations

	All	Odd	Even		Det	Sea
Factors						
Maintains Safe Classroom	.23** (n-114)		.35** (n=51)	.29** (n=64)	.36* (n=25)	
Encourages Safety	.30*** (n=97)	.17 (n=53)	.42** (n=44)	.49*** (n=56)		
Maintains Sanitary Classrom	.56*** (n=98)	.54*** (n=58)	.58** (n=40)	.38** (n=52)	.88*** (n=27)	
Provides Gross Motor Toys	.55*** (n=65)	.49*** (n=60)	.63*** (n=48)	.40*** (n=64)	.58*** (n=24)	.84*** (n=20)
Encourages Active Play	.25** (n=105)	.36** (n=57)	.12	.03	.62*** (n=27)	.46** (n=19)
Encourages Cognitive/language Dev.	.37*** (n=91)	.26* (n=48)	.49*** (n=43)	.57*** (n=55)		
Provides Creative Materials	.55*** (n=61)	.53*** (n=60)	.58*** (n=48)	.39*** (n=49)		.63*** (n=22)
Encourages Good Self-Concept	.36*** (n=112)	.33** (n=62)	.41** (n=50)	.42*** (n=61)		.58** (n=24)
Encourages Self-Help	.46*** (n=106)	.45** (n=27)	.49*** (n=49)	.40*** (n=62)	.40* (n=27)	.56** (n=17)
Encourages Social Behavior	.19* (n=100)		.28* (n=45)	.22* (n=22)		
Manages Class Activities Well	.31*** (n= 93)	.23* (n=53)	.43** (n=40)	.28* (n=53)	.30^{+} (n=25)	.19
Arranges Classroom Well	.49*** (n=112)	.34** (n=63)	.63*** (n=99)	.47*** (n=64)	.21 (n=27)	.78*** (n=21)
Constructs From Factors						
Environment	.42*** (n=93)	.41*** (n=55)	.42** (n=38)	.27* (n=49)	.86*** (n=25)	
Resources	.65*** (n=94)	.55*** (n=53)	.77*** (n=41)	.53*** (n=54)	.51** (n=24)	.91*** (n=16)
Management	.36*** (n=83)	.31** (n=48)	.43** (n=35)	.38** (n=49)	.32^{+} (n=25)	
Child Orientation	.37*** (n=82)	.20^{+} (n=43)	.50*** (n=39)	.51*** (n=51)		-.59 (n=5)
CDA Score	.46*** (n=49)	.35* (n=30)	.62** (n=19)	.36* (n=26)	.61** (n=21)	

+p<.10
*p<.05
**p<.01
***p<.001

Only those correlations significant at p<.15 are reported.

the CDA factors are sufficiently stable across days to be used in further analyses, although they are not as reliable as the AFI variables or the cognitive gain scores.

A variance components analysis also was undertaken to investigate the reliability of the twelve first-order factors and to determine the most appropriate unit of analysis. The variance components analysis was limited for several reasons:

- The analysis used only those classes in which the same caregiver was observed on two different days.
- Variance components analysis is difficult to perform if there are missing data; therefore only those caregivers for whom complete data were available were included in the analysis.
- The two factors that constituted the ENVIRONMENT construct had little variance and consequently were excluded from the analysis.
- For any classroom, observer and day of observation were confounded in the design. Two analysis approaches were employed to minimize contamination of variance component estimates by observer effects; nonetheless, the effects of observer and occasion cannot be separated.

The final sample available for the analyses contained only 52 of 133 classrooms. To examine the representativeness of the smaller sample, two validation analyses were undertaken. First, race-of-observer means were computed for each factor in each sample; no obvious differences were noted. Second, the twelve factors were factor-analyzed to redevelop second-order factors; similar factors emerged for both samples. Given these results, it is reasonable to assume that the two samples are not markedly different.

The results of the variance components analysis indicated that the occasion/observer is the predominant source of variance for the CDA Checklist. As noted above, the data collection design does not allow the effects of those two variance sources to be separated. Similar design problems also make it impossible to estimate the class and center components of variance accurately. However, since approximations of class and center components of variance showed them to be nearly equal, analyses were performed at the classroom level to parallel AFI and CFI process analyses and at the center level to confirm congruence.

The fact that estimated CDA reliabilities are not as high as those for other NDCS instruments is important primarily from the perspective of statistical power--low reliabilities reduce ability to detect actual effects. Although center-level analyses may be somewhat more reliable, the advantage of such analyses may be offset by the reduction in statistical power due to having fewer cases. Thus there is no strong reason to prefer one level of analysis over the other. It should also be noted that class-level CDA measures are of approximately the same reliability as class-level Child Focus measures, though both are less reliable than staff-level Adult Focus measures.

CHAPTER THREE: THE CDA ANALYSES

The CDA data collection effort was initially undertaken to provide information that might establish relationships between checklist items and the organizational framework behind the CDA credential. Given the existence of intuitively meaningful factors and adequate reliability, the usefulness of the checklist was further tested by including CDA rating variables in several input-process-outcome analyses.

Three questions were addressed in these CDA analyses:

- Are CDA variables influenced by the major independent variables (staff/child ratio, group size, number of caregivers, caregiver qualifications variables)?
- Are CDA ratings associated with the adult-focus and child-focus process variables?
- Are CDA ratings related to child gain scores?

To simplify interpretation, analyses are reported only for the CDA second-order factors ENVIRONMENT, RESOURCES, CLASS MANAGEMENT and CHILD ORIENTATION and for the overall CDA rating. Analyses were undertaken at both the classroom and the center level, with similar results. Results presented here were obtained at the classroom level.

Main Effects Analyses

The main effects analyses determined whether the CDA variables were influenced by the major independent (policy) variables. The policy variables, which are regulatable center characteristics, are:

- NUMBER OF CAREGIVERS: the total number of caregivers observed in each classroom;

- GROUP SIZE: the total number of children observed in a classroom or with a principally responsible caregiver;

- STAFF/CHILD RATIO: number of caregivers divided by group size; and

- CAREGIVER QUALIFICATIONS: described by four variables--total YEARS OF EDUCATION, presence or absence of specialized preparation related to young children (SPECIALIZATION), previous DAY CARE EXPERIENCE (prior to current job) and CENTER EXPERIENCE (tenure at current job).

Analytic methods included both correlation analysis and regression model development.

Correlations between the overall CDA score, four CDA constructs and the seven primary policy variables show a discernible pattern of relationships (Table 3.1). The two constructs that concern the physical classroom environment--ENVIRONMENT and RESOURCES--are associated with classroom structure as measured by both NUMBER OF CAREGIVERS and STAFF/CHILD RATIO. GROUP SIZE, however, which was found to be the most consistent predictor of many child (CFI) and adult (AFI) behaviors as well as test gain scores, was not highly correlated with the CDA variables. This variation in the process measures is partially explained if one considers differences between the measures. Both the CFI and the AFI reflect the proportion of time spent in various activities. There appears to be little doubt that caregivers and children interact more frequently in small groups. The CDA variables, however, are concerned with specific skills that are observed rather than with how often certain activities occur. It is reasonable to assume that teaching style is more a product of caregiver background and/or personality: caregivers who

Table 3.1

Correlation Matrix of CDA Variables with Major Policy Variables (n=118-135)

	CDA Score	Resources	Orientation	Management	Envir.	Group Size	Staff Count	Staff/ Child Ratio	Years of Ed.	Center Exp.	Prev. D.C. Exp.	Specification
CDA Score	1											
Resources	.64**	1										
Child Orientation	.74**	.35**	1									
Management	.74**	.24**	.39**	1								
Environment	.34**				1							
Group Size				-.12		1						
Staff Count	.24**	.39**			.15*	.52**	1					
Staff/Child Ratio	.20*	.30**			.14	-.42**	.50**	1				
Years of Educ.	.24**	.17*	.18*	.31**		-.19*		.20*	1			
Center Exp.		-.12			.12	-.11		.15	-.11	1		
Prev. D.C. Experience	.19*		.21**	.19*		-.12	.14	.30**	.11		1	
Specification	.16*		.20*	.23**					.34**	.27**	.23**	1

+p<.10
*p<.05
*p<.01
*p<.001

have developed good ways of teaching will exhibit this style regardless of the number of children involved. However, the proportion of time that she has the opportunity to evidence good skills may well be reduced in larger groups.

In general, the staff qualifications variables--YEARS OF EDUCATION, SPECIALIZATION and PREVIOUS DAY CARE EXPERIENCE--were not as strongly related to the classroom environment (ENVIRONMENT and RESOURCES). They were, however, positively associated with caregiver activity (CLASS MANAGEMENT and CHILD ORIENTATION). This provides evidence that more highly qualified caregivers show more interactive and better organized classroom behaviors. The general pattern of findings for the four constructs is reflected in the overall CDA SCORE, which shows significant positive correlations with five of the seven policy variables: NUMBER OF CAREGIVERS, STAFF/CHILD RATIO, YEARS OF EDUCATION, SPECIALIZATION and PREVIOUS DAY CARE EXPERIENCE.

On the basis of the correlations presented in Table 3.1, regression models were developed to predict the CDA constructs and total CDA rating using the policy variables. Because of the high correlations among the classroom structure policy variables and between SPECIALIZATION and the remaining qualifications policy variables, particular care was taken to construct simple models and avoid multicollinearity. The regressions, which were performed at both classroom and center level, indicated that the classroom composition variables NUMBER OF CAREGIVERS (class level) or STAFF/CHILD RATIO (center level) were the best predictors of the physical environment (ENVIRONMENT and RESOURCES), while the formal education variables YEARS OF EDUCATION and SPECIALIZATION (class level) were significant predictors of teaching skill and behaviors (CHILD ORIENTATION, CLASS MANAGEMENT). PREVIOUS DAY CARE EXPERIENCE also was found to predict CHILD ORIENTATION. A summary of results is presented in Table 3.2; details of the model development are contained in Tables C-1 to C-5 in Appendix C.

Table 3.2

Selected CDA Regression Results with Major Policy Variables (n=123)

Dependent Variable	Independent Variable	Simple Correlation	Regression Coefficient	Standard Error	F	R^2
CDA Score	Ratio	.20	.48	.22	4.64	.04
	Staff Count	.24	.05	.02	7.56	
	Yrs. Ed.	.24	.02	.01	7.30	.11
	Staff Count	.24	.04	.02	6.12	
	Prev. D.C. Exp.	.19	.02	.01	3.21	.08
Resources	Ratio	.30	.99	.29	11.94	.09
	Staff Count	.39	.10	.62	22.68	
	Yrs. Ed.	.17	.02	.61	5.07	.12
Child Orientation	Prev. Exp.	.21	.03	.02	3.72	.07
	Spec.	.20	.07	.04	3.12	
	Prev. D.C. Exp.	.21	.03	.01	4.66	.07
	Yrs. Ed.	.18	.02	.01	3.07	
Class Management	Yrs. Ed.	.31	.04	.01	12.73	.10
	Spec.	.23	.10	.04	6.57	.05
	Prev. D.C. Exp.	.19	.03	.02	4.31	.04
Environment	Staff Count	.15	.03	.02	2.83	.02

Classroom Process Analyses

The major focus of the classroom process analyses was to explore the relationships between the CDA variables and caregivers' (AFI) and children's (CFI) behaviors. Since the CDA Checklist was recorded immediately following an Adult Focus observation, it was expected that correlations between CDA and AFI variables would be especially strong. Patterns in relationships, however, might provide additional insights to the overall caregiver behavior dimension and ability of the CDA checklist to measure caregiver competence.

Adult Focus Analyses

The Adult-Focus Instrument includes a Physical Environment Inventory describing space, materials and equipment in the classroom; a Classroom Snapshot, which describes general activity patterns at a point in time; and a Five-Minute Interaction record, which describes the behavior of a particular caregiver in detail (Goodson, 1978). Eighteen variables and constructs from the Five-Minute Interaction were included in the analyses (See Appendix D).

A number of strong relationships were found between the CDA and AFI variables (see Table 3.3). CLASS MANAGEMENT, CHILD ORIENTATION and CDA SCORE showed particularly strong positive correlations with the AFI construct SOCIAL INTERACTION, in particular, DIRECT QUESTIONS and PRAISES, and negative correlations with OBSERVES and ADULT ACTIVITY. In other words, caregivers with higher ratings on these variables interacted with children more often and observed children or interacted with other adults less often. Caregivers with high CDA ratings also tended to focus less on the environment or large groups and more on small groups. However, the classroom environment variables

Table 3.3

Correlations of CDA Variables with Teacher Focus Variables

	ENVIRON-MENT (n=129)	RESOURCES (n=129)	CLASS MANAGE-MENT (n=124)	CHILD ORIEN-TATION (n=129)	CDA SCORE (n=124)
Commands			.12	.15	
Direct Questions	.21		.32	.44	.28
Responds	.20	.13	.28	.24	.27
Instructs	.18		.12	.19	.11
Adult Activity	-.16	-.16	-.44	-.34	-.37
Comforts	-.11		.26	.17	.17
Praises	.21	.12	.46	.48	.42
Corrects		-.15		-.10	-.12
Observes	-.19	-.13	-.25	-.17	-.23
Management Behavior		-.14			
Social Interaction	.26		.42	.47	.36
Focus to Environment		.11	-.26	-.27	-.13
Focus to Small Group			.36	.20	.21
Focus to Medium Group					
Focus to Large Group	.12	-.12	-.25		-.16

.20 = p<.01
.15 = p<.05
.11 = p<.10
.09 = p<.15

RESOURCES and ENVIRONMENT were not as strongly related to the AFI behaviors.

It should be noted that the CDA CLASS MANAGEMENT variable and the AFI MANAGEMENT construct were not strongly related. This is a reasonable result if one considers the constitution of the two variables. CLASS MANAGEMENT incorporates items that reflect positive ways in which the caregiver organizes classroom activities (i.e., make plans for day clear; announce time to clean up; provide five-minute warning before change in activity). On the other hand, the MANAGEMENT construct comprises two variables--command and correct--both of which are often associated with negative behavior and neither of which addresses classroom organization.

Child Focus Analyses

The Child-Focus Instrument describes children's behaviors at the end of twelve-second intervals over a twenty-minute period. Each child was observed for a total of sixty minutes over three mornings. The observations, however, were not performed on the same days as were the AFI/CDA observations. Results were aggregated to the classroom level (Connell, 1978). The sixteen variables that were included in the CFI/CDA analyses are described in Appendix D.

In general, the correlations between the CDA factors and the Child-Focus variables were stronger for teacher-directed activities than for free-play activities. Children with caregivers who received higher CDA ratings were more involved in constructive activities: they were less often seen as monitoring the environment, directing attention to children or wandering, and more often seen as receiving information and directing attention to adults. During free-play activities, caregivers with higher CDA ratings, especially for CHILD ORIENTATION, CLASS MANAGEMENT and CDA SCORE, were associated with children who less frequently wandered. The CDA/CFI correlations are reported in Tables C-6 to C-8, Appendix C.

The relationship between CDA and CFI variables is similar to that between AFI and CFI variables. Both sets of analyses indicate that caregivers who interact well with children are associated with children who are more involved. The overall classroom process analyses also suggest that the CDA Checklist captures many of the same aspects of caregiver behavior as the Adult Focus Instrument and therefore may be considered as an alternative measure of that dimension of classroom process.

Child Test Score Analyses

Analyses also were undertaken to determine the extent to which CDA behaviors influence children's generalized cognitive gain scores. Two outcome measures, the Preschool Inventory (PSI) and the Peabody Picture Vocabulary Test (PPVT), were used in the analysis (R. Goodrich and J. Singer, 1978). The investigation included the computation of correlations and the development of regression models using both the overall CDA rating and the four CDA constructs separately (see Tables 3.4 and 3.5). Regressions were weighted by the number of children tested in each classroom.

It should be noted that the process-outcome analyses (Singer, 1978) indicated that two AFI variables--SOCIAL INTERACTION and TO CHILD--also were highly correlated with PPVT GAIN. Since these variables were found to be strongly associated with CLASS MANAGEMENT and CHILD ORIENTATION, CDA findings are not surprising. Clearly, the two caregiver

Table 3.4

Regression Model Development for PPVT Change Scores
(n=119)

PPVT CHANGE SCORES	P	B	SE_B	F	R^2	P_f
CDA Rating Caregivers	.35	10.78	2.72	15.71	.12	.01
Constructs (n=124)						
Environment	.11	10.08	3.81	1.42	.011	.23
Resources	.26	6.00	2.02	8.85	.068	.01
Class Management	.36	8.82	2.04	18.69	.13	.01
Child Orientation	.30	7.01	1.99	12.48	.093	.01

Table 3.5

Regression Model Development for PSI Change Scores
(n=119)

PSI CHANGE SCORES	P	B	SE_B	F	R^2	P_f
CDA Rating Caregivers	.20	3.19	1.42	5.04	.041	.03
Constructs (n=123)						
Environment	-.02	-.81	1.89	0.18	.002	.67
Resources	.15	1.74	1.05	2.79	.023	.09
Class Management	.19	2.31	1.07	4.70	.037	.03
Child Orientation	.20	2.29	1.02	5.07	.040	.02

observation instruments both capture the dimension of caregiver behavior that is directed toward involvement with children, and the evidence is that this dimension is strongly related to children's cognitive development, particularly as measured by the PPVT.

CDA SCORE, RESOURCES, CLASS MANAGEMENT and CHILD ORIENTATION were found to predict both PSI and PPVT gain scores. The relationships, however, were considerably stronger for the PPVT.

The overall results also suggest that the PSF and the PPVT reflect different aspects of child development. GROUP SIZE is an especially strong predictor of PSI gain scores--much stronger than any of the staff background or caregiver behavior variables, although GROUP SIZE clearly affects the amount of interactive behavior in the classroom. On the other hand, PPVT gain scores appear to be less affected by classroom composition and more affected by type of caregiver. The quality of a caregiver is reflected both in background variables (especially SPECIALIZATION) and in classroom behaviors as measured by the CLASS MANAGEMENT and CHILD ORIENTATION variables. Thus there is evidence that caregivers who have specialized in a child-related field tend to exhibit more of the behaviors identified in the CDA philosophy, and that both these aspects of caregiver quality are associated with one dimension of cognitive growth.

CHAPTER FOUR: CONCLUSIONS AND RECOMMENDATIONS

The CDA Checklist, based on eleven of the functional areas defined by the CDA Consortium, is an additional measure of caregivers' skills and behaviors. It provides evidence that favorable classroom structure (NUMBER OF CAREGIVERS, STAFF/CHILD RATIO) is associated with day care centers that are safe, sanitary, and have many resources available and accessible to children. Caregiver qualifications, especially YEARS OF EDUCATION and/or SPECIALIZATION, are related to caregiver activities that are oriented toward child development and toward maintaining a well-organized classroom. Caregivers who direct their attention to children are associated with children who are more involved and who show greater gains in cognitive development, especially as measured by the PPVT. The CDA variables are strongly correlated with several of the Adult-Focus variables. The results of both teacher-focus analyses strongly indicate that what caregivers do in the classroom is associated with children's development.

The CDA Checklist has been developed and used only in the National Day Care Study. Although it clearly captures an important aspect of caregivers' skills and behaviors, it was not the intent of this study to test the instrument as a measure of caregiver competence. Before the instrument is used this way, the psychometric properties of the instrument must be established.

The above findings suggest that an important aspect of quality child care may be regulatable. Since the CDA Checklist is not directly related to the CDA credentialing process, an investigation of the relationships between CDA-credentialed caregivers and child outcomes should be undertaken before further recommendations are made. The checklist may be viewed as a potentially useful method of assessing caregivers at both the individual and the program level and of providing formative feedback to caregivers to help them strengthen their caregiving skills in all areas. It may also prove to be a useful way of standardizing the CDA credentialing process.

REFERENCES

Connell, D. "Child Focused Observations," in *Technical Appendices to the National Day Care Study*, Volume IV-C, Abt Associates Inc., 1978.

Goodrich, R. and J. Singer, "Cognitive Effects," in *Technical Appendices to the National Day Care Study*, Volume IV-C, Abt Associates Inc., 1978.

Goodson, B. "Adult Focused Observations," in *Technical Appendices to the National Day Care Study*, Volume IV-C, Abt Associates Inc., 1978.

Singer, J. " Process-Outcome Analyses," in *Technical Appendices to the National Day Care Study*, Volume IV-C, Abt Associates Inc., 1978.

Singer, J., Affholter, D., and Goodrich, R. "Variance Components and Reliability in the National Day Care Study," Abt Associates Inc., 1978.

APPENDIX A: CDA CHECKLIST

Directions: At the end of each day after observations are completed, stay in the center and answer all items in the following sections by circling the appropriate number indicating whether or not these situations occurred or materials were present.

There are 24 items on this instrument that could be answered "YES", "NO", or "NA", (Not Applicable or No Opportunity).

Code "YES" if the event occurred and the caregiver acted appropriately.

Code "No" if the event occurred but the caregiver did not act upon it. "No" is coded when the item has a potential to occur and does not.

Code "NA" if the event did not occur and thus there was no opportunity to take action. "NA" is coded when the item has no potential to occur. Example A) In question 1.1, if there were no spills, the caregiver does not have the potential to wipe them up; code "NA." Similarly, in item 1.11, if there were no rugs, code "NA." Example B) In item 9.11, if a child misbehaved by hitting another child and the caregiver was not angry or punitive, code "NO." If no such major misbehavior occurred, then code "NA."

For items regarding materials, observers should ask the caregiver to show them materials or equipment that are not visible in the center. You may only ask about items regarding materials. Do not ask about items regarding behaviors such as "Did the caregiver praise children for sitting quietly?" You must observe if the caregiver did this during the time you were in the center. If the behavior occurred code "YES;" if it did not occur while you were there code "No," or where applicable, "NA."

Developed by

Kathy Pope
Jane Stallings

April 1978

	1.0 SAFETY	Yes	No	
1.1	Did the caregiver clean up spills promptly (liquid, food)?	1	0	NA
1.2	Did the caregiver discuss safety rules?	1	0	
1.3	Did the caregiver enforce safety rules (i.e., no running down stairs, pushing, etc.)?	1	0	
1.4	Was outdoor equipment in safe condition?	1	0	
1.5	Was indoor furniture in safe condition?	1	0	
1.6	Did most furniture and equipment have rounded corners?	1	0	
1.7	Were floors and steps free of defects (loose, broken tiles, boards, etc.)?	1	0	
1.8	Were there barriers or railings around any porch, walkway, play area that are three feet or more above ground?	1	0	
1.9	Were all floors non-slippery when dry (e.g., not over-polished)?	1	0	NA
1.10	Were all toys and materials safe (free of danger)?	1	0	
1.11	Were all rugs rubber-backed, taped or nailed to the floor to prevent falls?	1	0	NA
1.12	Did all doors leading outside have exit signs?	1	0	
1.13	Were there any fire alarms (i.e., boxes) in the center?	1	0	
1.14	Was a fire extinguisher in the center?	1	0	
1.15	Were any evacuation signs posted in case of fire?	1	0	
1.16	Was parent information for most children available near the telephone?	1	0	NA
1.17	Was a community (or hospital) emergency number available near the telephone?	1	0	
1.18	Was physician's emergency number available near telephone?	1	0	

		Yes	No	
1.19	Was information on emergency first aid measures readily available (for burns, poisons)?	1	0	
1.20	Did all electrical outlets (within reach) have caps or covers?	1	0	
1.21	Were doors outside of center secured in a safe way (i.e., locks out of reach of children, but easily opened in emergencies)?	1	0	
1.22	Was storage out of reach of children for first aid supplies, medicines, cleaning agents, firearms, chemicals, cosmetics, etc?	1	0	
1.23	Were all well-traveled areas clear of material or equipment which could cause falls?	1	0	
1.24	Was access to all exits clear?	1	0	
1.25	Was yard fenced with safety gate(s) or lock(s)?	1	0	
1.26	Was a gate present at top and bottom of all stairs?	1	0	NA
1.27	Were railings (banisters) intact on all stairways?	1	0	NA
1.28	Were safety rules posted?	1	0	

2.0 HEALTH

		Yes	No	
2.1	Did the caregiver encourage or allow most children to wash their hands before eating?	1	0	
2.2	Did the caregiver wash hands before serving food?	1	0	
2.3	Did the caregiver talk with the children about how foods are prepared (e.g., applesauce, cookies, etc.)?	1	0	
2.4	Did the caregiver have any discussions with the children about health?	1	0	
2.5	Did the caregiver have any discussions with the children about health?	1	0	
2.6	Did the caregiver check children for colds, swollen glands, allergies, etc.?	1	0	
2.7	Was lighting adequate for children?	1	0	
2.8	Was the indoor temperature not excessively hot or cold for the children?	1	0	
2.9	Did the bathroom have soap?	1	0	
2.10	Were paper or cloth towels available for children?	1	0	
2.11	Were tissues (e.g., Kleenex, toilet tissue) available for the children?	1	0	
2.12	Were toothbrushes available for most of the children?	1	0	
	2.12.1 (IF YES,) Did most children brush their teeth at the center?	1	0	NA
	2.12.2 Were most toothbrushes labeled with children's names or identification markings?	1	0	NA
2.13	Were there provisions made for laundering clothes, blankets, cot covers?	1	0	
2.14	Was at least one menu posted?	1	0	
2.15	Were there any posters on the wall about nutritious/balanced foods?	1	0	
2.16	Were first aid suplies available and stored together (soap, antiseptics, bandages, thermometers, etc.)?	1	0	
2.17	Were there any posters on the walls about first aid (such as cleaning and bandaging wounds)?	1	0	

		Yes	No	
2.18	Was a list kept of children with allergies?	1	0	
2.19	Were child health records kept?	1	0	
2.20	Was there an isolated area for removal of sick children?	1	0	
2.21	Were all the hallways clean?	1	0	NA
2.22	Were all the indoor areas clean?	1	0	
2.23	Was the kitchen clean and free of odors?	1	0	
2.24	Was the bathroom clean and free of odors?	1	0	
2.25	Was the play yard clean and litter-free?	1	0	
2.26	Were all tables clean for snacks?	1	0	
2.27	Were there a clean rug for children to sit on?	1	0	
2.28	Was all garbage disposed of properly (food)?	1	0	
2.29	Was all trash disposed of properly (paper)?	1	0	
2.30	Was the food storage area free of insects or rodents?	1	0	
2.31	Was the food storage area clean and free of odors?	1	0	
2.32	Were cot covers clean?	1	0	
2.33	Were disposable dished and/or utensils used?	1	0	
	2.33.1 (I NO,) were all dishes washed hygenically (e.g., use of detergents or dishwasher)?	1	0	NA
2.34	Were all the rooms well ventilated (free of smoke, steam, etc.)	1	0	
2.35	Were all toys clean and free of foods and/or other sticky substances?	1	0	
2.36	Were there individual cups or glasses for drinking?	1	0	

3.0 PHYSICAL COMPETENCE

Below is a list of equipment. Please indicate whether each was present, and if it was used by the children. If it was not present, circle 0.

		PRESENT	NOT PRESENT
3.1	Tricyles	1	0
3.2	Jungle gyms, climbers	1	0
3.3	Ladders	1	0
3.4	Clmbing ropes	1	0
3.5	Balancing beams or bouncing boards	1	0
3.6	Basketball and hoop (child level)	1	0
3.7	Swings or slides	1	0
3.8	Wagons	1	0
3.9	Roller skates	1	0
3.10	Child sized work tools (brooms, rakes, dust pans)	1	0
3.11	Blocks	1	0
3.12	Small wheel toys, cars, trucks, etc.	1	0
3.13	Child size skill equipment: carpentry, sewing	1	0
3.14	Balls or bean bags	1	0
3.15	Hula hoops	1	0
3.16	Swimming or wading pool	1	0
3.17	Child-size chairs and tables	1	0

4.0 COGNITIVE AND LANGUAGE DEVELOPMENT

		YES	NO
4.1	Did the caregiver encourage discussions at snack and/or lunch?	1	0
4.2	Did the caregiver schedule any discussion groups	1	0
4.3	Did the caregiver read books with children?	1	0
4.3.1	Did the caregiver ask questions while reading books with children?	1	0
4.4	Did the caregiver speak to and listen to children at eye level?	1	0
4.5	Did the caregiver explain, inform, or discuss such topics as insects, food, weather, etc. (i.e., their environment, the world)?	1	0
4.6	Did the caregiver encourage the children to identify ingredients and foods at snack or lunchtime?	1	0
4.7	Did the caregiver ask questions of children and encourage them to speak during the discussions of general topics?	1	0
4.8	Did the caregiver encourage or allow any children to explore different aspects of water in relation to other materials (e.g., objects that float on water change water color, blow bubbles, paint with water on cement to watch it evaporate)?	1	0
4.9	Did the caregiver encourage most children to look at books some of the time?	1	0
4.10	Did the caregiver ask children to speak in complete sentences?	1	0
4.11	Did the caregiver encourage children to pronounce words clearly?	1	0
4.12	Were resource people brought in to work with language development (e.g., librarian, speech specialists, or others)?	1	0
4.13	Were several age-appropriate books available (i.e., more than 10)?	1	0
4.14	Were there materials or equipment for experimentation in science activities?	1	0

		YES	NO
4.14.1	(IF YES) Did the caregiver facilitate use of materials or equipment for experimentation in science activities?	1	0
4.15	Were there any language development teaching machines (e.g., language masters, records, audio tapes)?	1	0
4.15.1	(IF YES) Were any language development teaching machines used?	1	0

5.0 CREATIVE EXPRESSION

		YES	NO
5.1	Did the caregiver encourage or allow children to sing spontaneously?	1	0
5.2	Did the caregiver encourage or allow children to express their feelings and ideas?	1	0
5.3	Did the caregiver encourage or allow children to do art projects any way they liked?	1	0
5.4	Did the caregiver encourag or allow children to do such things as act out songs, poems or stories?	1	0
5.5	Did the caregiver encourage or allow children to dance freely?	1	0
5.6	Did the caregiver encourage or allow children to make up games?	1	0

Below is a list of materials. Please indicate whether each was present by circling 1. If it was not present, circle 0.

		PRESENT	NOT PRESENT
5.7	Doll houses	1	0
5.8	Dress up clothes	1	0
5.9	Dramatic play props	1	0
5.10	Dolls	1	0
5.11	Ethnic dolls	1	0
5.12	Art materials: pencils, crayons, paper, scissors paints, etc.	1	0
5.13	Records for dance or spontaneous expressions	1	0
5.14	Musical instruments	1	0
5.15	Puppets (paper, cloth, etc.)	1	0
5.16	Weaving materials	1	0
5.17	Doll house equipment, toy village, toy farm, etc.	1	0
5.18	Sand to be in (box or area)	1	0

		PRESENT	NOT PRESENT
5.19	Dirt to dig in for planting	1	0
5.20	Materials to explore: modeling clay, play don, finger paints, clay, mud, glue, starch, etc.	1	0
5.21	Nature materials: rocks, feathers, fur, leather plants shells	1	0
5.22	Animals to hold	1	0
5.23	Animals to observe: birds, fish, caged animals	1	0

6.0 SELF-CONCEPT

		YES	NO
6.1	Did the caregiver allow any children to create their own picture books, greeting cards, etc.?	1	0
6.2	Did the caregiver acknowledge children who had been absent, and say things like "We missed you, we are glad you're back?"	1	0
6.3	Did the caregiver use songs, exercises, etc. that use the names of children?	1	0
6.4	Did the caregiver sue songs, exercises, etc. that direct children to identify parts of their bodies?	1	0
6.5	Did the caregiver ask about any of the children's families?	1	0
6.6	Did the caregiver acknowledge most children by their names?	1	0
6.7	Did the caregiver keep from comparing children in unfavorable ways?	1	0
6.8	Were any children told it is all right to make mistakes sometimes?	1	0
6.9	Did the caregiver praise children by name for correct answers or good performance?	1	0
6.10	Did the caregiver ask children to assist in tasks?	1	0
6.11	Were charts posted that show heights and weights of most children by name?	1	0
6.12	Were there any pictures or posters on walls that reflect ethnic groups?.	1	0
6.13	Were any objects present that reflect various holiday customs or cultures (i.e., Pinatas, Hanukkah candles, food)?	1	0
6.14	Were tape recorders available?		
	6.14.1 (IF YES) Were tapes made of children's voices, and played back for them to hear?	1	0
6.15	Were any of the children's self portraits, (paintings or clay) displayed?	1	0
6.16	Were any photographs of children displayed?	1	0

		YES	NO
6.17	Were there name tages or symbols for the children on the place for their personal belongings (e.g., cubbies, coat racks, or other places)?	1	0
6.18	Were any mirrors provided for children to see themselves (i.e., full length, or child-level)?	1	0

7.0 SELF-HELP/INDIVIDUAL STRENGTH

		YES	NO
7.1	Did the caregiver encourage or allow children to carry their own trays or dishes?	1	0
7.2	Did the caregiver encourage or allow children to help put trash in baskets or take trash out to cans?	1	0
7.3	Did the caregiver encourage or allow children to be responsible for hanging up outdoor clothes?	1	0
7.4	Did the caregiver encourage or allow children to clean up their own toys and put things away when finished?	1	0
7.5	Did the caregiver encourage or allow children to fold up their own blanket after nap?	1	0
7.6	Did the caregiver encourage or allow chidlren to put up their own coats?	1	0
7.7	Did the caregiver encourage or allwo children to tie or buckle their shoes?	1	0
7.8	Did the caregiver encourage children to zip or button their own clothes?	1	0
7.9	Did the caregiver allow most children to choose some of their own indoor or outdoor activities?	1	0
7.10	Did the caregiver encourage or allow children to wash their own hands?	1	0
7.11	Did the caregiver encourage or allow children to pour their own drinks?	1	0
7.12	Did the caregiver encourage or allow children to pass food to other children?	1	0
7.13	Did the caregiver encourage or allow children to blow their own noses?	1	0

8.0 PROSOCIAL BEHAVIOR

		YES	NO
8.1	Did the caregiver ask children to listen to each other during group times?	1	0
8.2	Did the caregiver sometimes allow (or give the opportunity) for children to help others?	1	0
8.3	Did the caregiver sometimes request children to share?	1	0
8.4	Did the caregiver sometimes praise children for sharing?	1	0
8.5	Did the caregiver thank children for helping?	1	0
8.6	Did the caregiver ask children to consider each others feelings?	1	0
8.7	Did the caregiver encourage children to take turns in using materials (i.e., riding tricycles, etc.)?	1	0

9.0 CAREGIVER ORGANIZATIONAL QUALITIES (Management Skills)

		YES	NO
9.1	Were most learning materials accessible (conveniently within reach) to the children?	1	0
9.2	Overall, did the environment seem well-arranged (orderly)?	1	0
9.3	Were special play areas defined (block corner, quiet area, dress up, etc.)?	1	0
9.4	Did the caregiver sometimes work with a small group of children?	1	0
9.5	Were any quiet activities arranged or available for the children?	1	0
9.6	Were any active activities arranged or available for the children?	1	0
9.7	Were any groups arranged or available where children worked together on a project?	1	0
9.8	Were any groups arranged or available where most children listened to each other or to a story?	1	0
9.9	Did the caregiver usually wait for children to settle down, or praise or acknowledge children who were sitting quietly when she wanted their attention?	1	0
9.10	Was the caregiver free from anger and punitiveness when moderate misbehavior occurred?	1	0
9.11	Was the caregiver free from anger and punitiveness when major misbehavior occurred?	1	0
9.12	Did the caregiver ever discuss rules with children?	1	0
9.13	When a child broke a rule, did the caregiver usually make clear in a positive way that he/she had done so?	1	0
9.14	Were there any rules for the number of chidlren allowed in each area?	1	0
9.15	Were plans for the day made clear for the children?	1	0
9.16	Did the caregiver usually announce to children (individually or group) when it was time to clean up?	1	0

		YES	NO
9.17	In changing activities, did the caregiver use methods such as "follow the leader" to guide children into the next activity?	1	0
9.18	Did the caregiver usually give directions in a clear positive way?	1	0
9.19	Did the caregiver use songs, rhymes, etc. to get children to help clean up?	1	0
9.20	Did the caregiver usually give a 5-minute (or so) warning time to children when an activity was ab out to change?	1	0
9.21	Did the caregiver remind children to go to the bathroom?	1	0
9.22	Were there enough cubbies for each child's personal belonging?	1	0
9.23	Was any of the children's art work displayed?	1	0
9.24	Were some bulletin boards at child-eye level?	1	0
9.25	Was there a science area?	1	0
9.26	Was there a listening area (records, tapes, language master)?	1	0
9.27	Was there a reading area or quiet area?	1	0
9.28	Were there any small, private areas where a child could be alone?	1	0

10.0 OBSERVER'S OVERALL JUDGMENT OF DAY CARE CENTER ENVIRONMENT

		YES	NO
10.1	Did children have a number of minor accidents (e.g., cuts, bumps, bruises)?	0	1
10.1.1	DESCRIBE: ____________________		
10.2	Was the physical environment pleasant?	1	0
10.2.1	DESCRIBE: ____________________		
10.3	Was the emotional environment pleasant, supportive personalized?	1	0
10.3.1	DESCRIBE: ____________________		
10.4	Did the caregiver show favoritism?	1	0
10.4.1	DESCRIBE: ____________________		
10.5	Did children often intrude upon each other's activities? (Argue or fight over materials or equipment)	1	0
10.5.1	DESCRIBE: ____________________		
10.6	Did the caregiver use a negative tone with children?	1	0
10.6.1	DESCRIBE: ____________________		

		YES	NO
10.7	Was the caregiver clear, firm and positive when giving guidelines to the children?	1	0
10.7.1	DESCRIBE: ______________________________		
10.8	Overall, did the center seem crowded?	1	0
10.8.1	DESCRIBE: ______________________________		
10. 9	Were there many accidental spills or breakage?	1	0
10.9.1	DESCRIBE: ______________________________		
10.10	Were there any activities that the caregiver relinquished to another adult? (puppet shows, water play, music, exercise, etc.)	1	0
10.10.1	DESCRIBE: ______________________________		
10.11	Was there time for living? (Not herded or hurried?)	1	0
10.11.1	DESCRIBE: ______________________________		
10.12	Does the caregiver hold, cuddle or display affection with children?	1	0
10.12.1	DESCRIBE: ______________________________		

		YES	NO
10.13	Is the classroom moderately noisy? (Describe whether it was too calm or too noisy and confused.)	1	0

10.13.1 DESCRIBE: ____

____

____

10.4	Is the caregiver flexible? (able to change plans as new events arise)	1	0

10.14.1 DESCRIBE: ____

____

____

		OFTEN	SOMETIMES	SELDOM
11.1	The caregiver is aware of dangerous situations.	1	2	3
11.2	The caregiver cautions children quickly and appropriately if there is physical danger to themselves or other children.	1	2	3
11.3	During active play periods the caregiver organizes games which involve running jumping, skipping, etc.	1	2	3
11.4	During active play periods the caregiver participates in games which involve running, jumping, skipping, etc.	1	2	3
11.5	The caregiver provides activities and experiences which encourage questioning, probing and problem-solving skills.	1	2	3
11.6	During activity times the caregiver provides a variety of experience and materials that stimulate the children to explore and express their creative abilities .	1	2	3
11.7	The caregiver is genuinely concerned about how the children feel.	1	2	3
11.8	The caregiver responds to individual needs in a way which encourages children to appreciate themselves.	1	2	3
11.9	The caregiver allows the children to pursue their own interests independently.	1	2	3
11.10	The caregiver encourages children to try new things on their own.	1	2	3
11.11	In free play situations the caregiver encourages children to play with one another.	1	2	3
11.12	The caregiver encourages children to respect the needs and feelings of other children.	1	2	3
11.13	The caregiver maintains order without yelling or addressing the children in a negative manner.	1	2	3

APPENDIX B: TWELVE PRELIMINARY FACTORS IDENTIFIED FOR CDA ANALYSIS

SAFETY

Safe Classroom

- Was indoor furniture in safe condition?
- Were all toys and materials safe?
- Was information on emergency first aid available?
- Were well-traveled areas clear?
- Was access to all exits clear?

Safety Awareness

- Did caregiver discuss safety rules?
- Did caregiver enforce safety rules?
- Was caregiver aware of dangerous situations?
- Did caregiver caution children quickly?

HEALTH

- Were the hallways clean?
- Were indoor areas clean?
- Was the bathroom clean?
- Was the play yard clean?
- Was garbage disposed of properly?
- Was trash disposed of properly?
- Were there insects or rodents in food storage area?
- Were toys clean?

PHYSICAL COMPETENCE

1. Gross Motor Toys Available

- Tricycles
- Jungle gyms
- Ladders
- Climbing ropes
- Balancing beams or bouncing boards
- Wagons
- Child-size work tools
- Small wheel toys
- Child-size skill equipment
- Balls or bean bags
- Hula hoops

2. Active Play

- Did caregiver organize active play?
- Did caregiver participate in active play?

COGNITIVE LANGUAGE DEVELOPMENT

- Did caregiver encourage discussions at meals?
- Did caregiver schedule discussion groups?
- Did caregiver encourage children to identify ingredients at meals?
- Did caregiver explain about environment?
- Did caregiver speak and listen at child eye level?
- Did caregiver ask questions during discussions?
- Did caregiver request complete sentences?
- Did caregiver encourage problem-solving skills?

CREATIVE EXPRESSION

Materials Present

- Doll houses
- Dress-up clothes
- Dramatic play props
- Dolls
- Ethnic dolls
- Musical instruments
- Puppets
- Weaving materials
- Doll house equipment
- Materials to explore

SELF CONCEPT

- Did caregiver ask about children's families?
- Did caregiver acknowledge children by name?
- Were children told mistakes are OK sometimes?
- Did caregiver praise children by name?
- Did caregiver ask children to assist?
- Caregiver was concerned about children's feelings.
- Caregiver encouraged children to appreciate themselves.

INDIVIDUAL STRENGTH

- Did caregiver allow children to:
  - Help with trash disposal?
  - Put toys away?
  - Pour own drinks?
  - Pass food to others?

PROSOCIAL BEHAVIOR

- Did caregiver ask children to listen to each other?
- Did caregiver request sharing?
- Did caregiver praise sharing?
- Did caregiver thank for helping?

ORGANIZATIONAL QUALITIES

1. Classroom Management

- Did caregiver work with small groups?
- Were quiet activities arranged?
- Were there groups for projects?
- Did caregiver wait for kids to settle down?
- Was caregiver free of anger with moderate misbehavior?
- Did caregiver discuss rules with children?
- Were there rules for area occupation?
- Were plans for the day made clear?
- Did caregiver announce time to clean up?
- Did caregiver guide to new activity?
- Did caregiver give clear, positive directions?
- Did caregiver use songs, etc., for cleanup?
- Did caregiver give 5-minute warnings?
- Did caregiver maintain order without yelling?

2. Classroom Organization

- Were learning materials accessible?
- Was environment well-arranged?
- Were special play areas defined?
- Was there a science area?
- Was there a listening area?
- Was there a reading/quiet area?
- Was there a private area?

APPENDIX C

Table C-1

CDA Rating Regression Model Development

CDA SCORE	P	B	SE	F	R^2	P_f
CLASS LEVEL (n=118)						
Number of Caregivers	.24	-.04	.017	7.06	.057	.10
Staff/Child Ratio	.20	.48	.22	4.64	.038	.03
Years of Education	.24	.02	.009	6.80	.055	.10
Specialization	.16	.048	.027	2.99	.025	.08
Previous Day Care Exp.	.18	.023	.012	4.11	.034	.04
Number of Caregivers	.24	.035	.019	3.46	.067	.06
Staff/Child Ratio	.20	.26	.25	1.13		
Number of Caregivers	.24	.045	.016	7.56	.114	.01
Years of Education	.24	.023	.008	7.30		.01
Number of Caregivers	.24	.044	.016	6.75	.079	.01
Specialization	.16	.044	.027	2.72		.10
Number of Caregivers	.24	.042	.017	6.12	.083	.01
Previous Day Care Exp.	.16	.020	.011	3.21		.07
Staff/Child Ratio	.20	.38	.22	2.95	.079	.09
Years of Education	.24	.020	.009	5.06		.02
Staff/Child Ratio	.20	.46	.22	4.36	.061	.04
Specialization	.16	.045	.027	2.73		.10
Years of Education	.24	.020	.009	4.66	.063	.03
Specialization	.16	.028	.029	0.95		.33
Years of Education	.24	.020	.008	5.52	.078	.02
Previous Day Care Exp.	.18	.019	.012	2.87		.09
CENTER LEVEL (n=53)						
Staff/Child Ratio	.41	.95	.30	10.09	.17	.01
Staff/Child Ratio	.41	1.25	.33	14.72	.23	.01
Center Experience	-.04	-.017	.0084	4.11		.04
Staff/Child Ratio	.41	1.19	.32	14.04	.30	.01
Center Experience	-.04	-.020	.0083	5.74		.02
Specialization	.28	.077	.037	4.28		.04

Table C-2

ENVIRONMENT Regression Model Development

ENVIRONMENT	P	B	SE	F	R^2	P_f
CLASS LEVEL (n=123)						
Number of Caregivers	.15	.026	.016	2.83	.023	.09
Staff/Child Ratio	.14	.32	.20	2.49	.020	.09
Center Experience	.12	.007	.005	1.82	.015	.18
Number of Caregivers	.15	.01	.018	1.14	.029	.29
Staff/Child Ratio	.14	.21	.23	0.80		.37
Number of Caregivers	.15	.025	.016	2.65	.036	.10
Center Experience	.12	.006	.005	1.65		.20
Staff/Child Ratio	.14	.30	.20	2.11	.032	.15
Center Experience	.12	.006	.005	1.46		.23
Number of Caregivers	.15	.019	.018	1.14	.041	.29
Staff/Child Ratio	.14	.18	.23	0.62		.43
Center Experience	.12	.006	.005	1.46		.23
CENTER LEVEL (n=53)						
Staff/Child Ratio	.26	.58	.30	3.83	.07	.05
Staff/Child Ratio	.26	.50	.30	2.86	.11	.09
Specialization	.24	.055	.038	2.08		.15

Table C-3

RESOURCES Regression Model Development

RESOURCES	P	B	SE	F	R^2	P_f
CLASS LEVEL (n=123)						
Number of Caregivers	.39	.098	.021	21.29	.15	.01
Staff/Child Ratio	.30	.99	.29	11.94	.090	.01
Years of Education	.17	.021	.011	3.729	.030	.05
Group Size	.02	.001	.003	.057	.000	.81
Number of Caregivers	.39	.080	.024	11.21	.17	.01
Staff/Child Ratio	.30	.50	.311	2.58		.11
Number of Caregivers	.39	.099	.021	22.68	.18	.01
Years of Education	.17	.023	.010	5.07		.02
Number of Caregivers	.39	.13	.024	29.21	.20	.01
Group Size	.02	-.009	.004	6.95		.01
Staff/Child Ratio	.30	.92	.28	10.18	.106	.01
Years of Education	.17	.016	.011	2.14		.14
Staff/Child Ratio	.30	1.26	.31	15.99		.01
Group Size	.02	.007	.004	3.84	.12	.05
Number of Caregivers	.39	.086	.024	13.01	.19	.01
Staff/Child Ratio	.30	.38	.31	1.43		.23
Years of Education	.17	.021	.010	3.88		.05
Staff/Child Ratio	.30	1.20	.31	14.72	.14	.01
Group Size	.02	.008	.004	4.59		.03
Years of Education	.17	.018	.011	2.89		.09
Number of Caregivers	.39	.13	.024	28.42	.22	.01
Group Size	.02	-.008	.004	5.31		.02
Years of Education	.17	.019	.010	3.48		.06
CENTER LEVEL (n=53)						
Number of Caregivers	.54	.13	.030	20.51	.29	.01
Number of Caregivers	.54	.11	.033	11.03	.32	.01
Staff/Child Ratio	.42	.66	.40	2.73		.10

Table C-4

CLASS MANAGEMENT regression Model Development

CLASS MANAGEMENT	P	B	SE	F	R^2	P_f
CLASS LEVEL (n=118)						
Years of Education	.31	.042	.012	12.73	.099	.01
Specialization	.23	.097	.038	6.57	.054	.01
Previous Day Care Exp.	.19	.034	.016	4.31	.036	.04
Years of Education	.13	.039	.012	10.96	.12	.01
Previous Day Care Exp.	.19	.026	.016	2.71		.10
Years of Education	.31	.036	.012	8.40	.12	.01
Specialization	.23	.061	.039	2.50		.11
Years of Education	.31	.035	.012	7.99		.01
Previous Day Care Exp.	.19	.021	.016	1.62		.20
Specialization	.23	.048	.040	1.40		.24
CENTER LEVEL (n=52)						
Specialization	.38	.14	.049	8.39	.14	.01
Specialization	.38	.13	.049	6.63	.20	.02
Staff/Child Ratio	.30	.68	.38	3.28		.07

Table C-5

CHILD ORIENTATION Regression Model Development

CHILD ORIENTATION	P	B	SE	F	R^2	P_f
CLASS LEVEL (n=123)						
Previous Day Care Exp.	.21	.034	.015	5.45	.043	.02
Specialization	.20	.081	.037	4.84	.038	.03
Years of Education	.18	.023	.011	3.84	.031	.05
Previous Day Care Exp.	.21	.029	.015	3.72	.067	.05
Specialization	.20	.066	.037	3.12		.08
Previous Day Care Exp.	.21	.032	.015	4.66	.067	.03
Years of Education	.18	.020	.011	3.07		.08
Specialization	.20	.064	.039	2.64	.052	.10
Years of Education	.18	.016	.012	1.67		.20
Previous Day Care Exp.	.21	.028	.015	3.57	.079	.06
Specialization	.20	.050	.040	1.58		.21
Years of Education	.18	.015	.012	1.53		
CENTER LEVEL (n=53)						
Specialization	.39	.14	.048	9.15	.15	.01
Specialization	.39	.15	.046	10.38	.23	.01
Number of Caregivers	.26	.065	.030	4.89		.03

Table C-6

Correlations of CDA Variables with Teacher-Directed CFI Variables

	ENVIRON-MENT (n=123)	RESOURCES (n=123)	CLASS MANAGE-MENT (n=119)	CHILD ORIEN-TATION (n=123)	CDA SCORE (n=119)
Monitors Environment	-.17	-.16	-.13	-.30	-.27
Wanders	-.14	-.21	-.14	-.22	-.24
Moves W/ Purpose			.15		
Gives Opinions	-.11	.11	-.18		.09
Gives Orders		-.11	-.12	-.22	-.19
Receives Orders					
Receives Info.	.20	.21	.18	.27	.30
Receives Genr'l	-.10		.11		
Attn to Adults	.13	.25	.22	.23	.29
Attn to Children	-.23	-.15	-.15	-.28	-.26
Attention to Environment		-.13	-.09		-.15
Attn to Groups	.11				
No Task	-.20			-.23	-.18
Open Activity			-.13		-.13
Structured Activity		.21		.12	.15
Considers	-.10	-.14			-.11

.20 = p<.01
.15 = p<.05
.11 = p<.10
.09 = p<.15

Table C-7

Correlations of CDA Variables with Free Play CFI Variables

	ENVIRONMENT (n=123)	RESOURCES (n=123)	CLASS MANAGEMENT (n=119)	CHILD ORIENTATION (n=123)	CDA SCORE (n=119)
Monitors Environment	-.16			-.15	-.14
Wanders	-.16	-.13	-.22	-.22	-.22
Gives Opinions	-.14				
Gives Orders		.10			
Receives Orders			-.18		-.09
Receives Info.	.12		.13	.15	.12
Receives Genr'l	-.16	-.09		-.11	-.11
Attn to Adults	.09				
Attn to Children	-.25			-.18	-.13
Attention to Environment	.15	.12		.10	-.11
Attn to Groups		-.16			
No Task	-.13	.11			
Open Activity		.23			.12
Structured Activity	.12	-.18		.09	
Considers					

.20 = p<.01
.15 = p<.05
.11 = p<.10
.09 = p<.15

Table C-8

Correlations of CDA Variables with Comb. (F.P+T.D) CFI Variables

	ENVIRONMENT (n=126)	RESOURCES (n=126)	CLASS MANAGEMENT (n=122)	CHILD ORIENTATION (n=126)	CDA SCORE (n=122)
Monitors Environment	-.14	-.15	-.12	-.27	-.25
Wanders	-.13	-.20	-.23	-.26	-.27
Moves W/ Purpose	.11		.21	.12	.13
Gives Opinions	-.12		.15		
Gives Orders					
Receives Orders					
Receives Info.	.16	.16		.20	.21
Receives Genr'l	-.10	-.09			
Attn to Adults		.12		.11	.14
Attn to Children	-.23			-.23	-.16
Attention to Environment					
Attn to Groups		-.17			
No Task	-.19			-.17	
Open Activity		-.21			
Structured Activity			-.09	.15	
Considers					

.20 = $p<.01$
.15 = $p<.05$
.11 = $p<.10$
.09 = $p<.15$

198

APPENDIX D: VARIABLES INCLUDED IN ANALYSIS

Adult Focus Instrument

COMMANDS	Self-explanatory.
CORRECTS	Self-explanatory.
DIRECT Q.	Proportion of time teacher poses a direct question, e.g., "What is your favorite color?"
RESPONDS	Self-explanatory.
INSTRUCTS	Self-explanatory.
ADULT ACT.	Proportion of time teacher engages in self-related activity or conversation with other adults.
COMFORTS	Self-explanatory.
PRAISES	Self-explanatory.
OBSERVES	Proportion of time teacher spends listening to or observing others.
MANAGEMENT BEHAVIOR	Commands and corrects.
SOCIAL INTERACTION	Direct Q and responds and instructs and comforts and praises.
FOCUS TO ENVIR.	Proportion of time teacher focuses attention to environment rather than to children.
FOCUS TO SMALL GROUP	Proportion of time teacher focuses attention to a small group--defined as 2-7 children.
FOCUS TO MEDIUM GROUP	Proportion of time teacher focuses attention to a medium group--defined as 8-12 children.
FOCUS TO LARGE GROUP	Proportion of time teacher focuses attention to a large group--defined as 13 or more children.

Child Focused Instrument

MONITORS ENVIRONMENT	Proportion of time spent monitoring environment; child's attention is obviously directed at other people or things.
WANDERS	Proportion of time spent wandering around center with no apparent purpose to his/her movement. Child may be sitting or standing doing nothing, looking around the area with no apparent focus.
MOVES WITH PURPOSE	Proportion of time child moves with purpose; child is going from one activity to another; evident that there is some goal to movement.
GIVES OPINIONS	Proportion of time child gives opinions, comments, information or states preferences.
GIVES ORDERS	
RECEIVES ORDERS	Proportion of time child receives commands with which compliance is expected.
RECEIVES INFORMATION	Proportion of time child receives instruction, materials or assistance related to a task or problem.
RECEIVES GEN'L	Proportion of time child is asked for intermation or receives commends of a general nature.
ATTENTION TO ADULTS	Proportion of time child focuses attention on caregivers or other adults.
ATTENTION TO CHILDREN	Proportion of time child focuses attention on other children.

ATTENTION TO ENVIRONMENT	Proportion of time child focuses attention on something other than caregivers or children.
ATTENTION TO GROUPS	Proportion of time child focuses attention on a group of children or caregivers.
NO TASK	Proportion of time spent in no apparent task or activity.
OPEN ACTIVITY	Proportion of time spent in open-ended, expressive activity.
STRUCTURED ACTIVITY	Proportion of time spent in closed, structure activity.
CONSIDERS	Proportion of time spent considering, contemplating, tinkering; e.g., child struggles with a problem attempting to solve it.

Interviews with Parents

Jean Layzer

TABLE OF CONTENTS

Page

CHAPTER ONE: INTRODUCTION 187

CHAPTER TWO: SUMMARY OF THE PHASE II PARENT INTERVIEW FINDINGS 189

Parent Expectations

Parent Involvement 190

Parent Satisfaction 190

Conclusions 190

CHAPTER THREE: THE PHASE III SAMPLE 192

CHAPTER FOUR: PHASE III PARENT INTERVIEW FINDINGS 195

Parents' Goals for Children in Day Care 195

General Day Care Preferences 197

Preferences Concerning the Regulatory Variables 202

Parents' Knowledge of the Child's Day 207

Parent-Center Communications 210

Parent Involvement 213

Parents' Use of Center Services 214

Conclusions 215

CHAPTER ONE: INTRODUCTION

Interviews were conducted with parents of children who participated in the study during Phases II (1975-1976) and III (1976-1977) of the NDCS. Although the two interviews dealt with some common issues, they were different in scope and served very different purposes. The Phase II Parent Interview, a survey of 1165 parents, was an integral part of the overall study and provided several kinds of important information. First, the interview served to introduce the study to parents, to obtain permission for their child's participation and to answer their questions about study procedures. Thus, it was essential to try to interview all parents rather than a sample.

Parents were asked to provide basic demographic information as well as information on their attitudes toward and practices regarding the raising of children and the family's previous experience with day care. Because these background factors can influence both child behavior and test performance, they must be taken into account in any assessment of the effects of center characteristics on children.

In addition, parents' opinions were solicited on a number of topics. The reasons for their choice of center care rather than other kinds of care were examined. Parents were asked about the extent to which they were involved in center activities and the nature of that involvement; their expectations of the center; and how satisfied they were with their particular center. Finally, parents whose children received federally-subsidized care were asked about their use of available social and health services.

In Phase III, in-depth interviews were conducted with 112 parents distributed across 12 publicly supported study centers (four in each site). Parents selected for participation represented those families most affected by federal day care policy--lower income, urban, minority, single-parent families. Information was collected on parental views on the importance of different policy variables, parent goals for children in day care, parent involvement in day care and family use of center services.

205

CHAPTER TWO: SUMMARY OF THE PHASE II PARENT INTERVIEW FINDINGS

In the fall of 1975, interviews were conducted with 1165 parents of target children in 64 centers. In all cases, only one parent was interviewed; in almost every instance this was the mother of the target child. All parents provided the basic family background data needed for the effects analyses. The survey obtained information on background characteristics of families, parental attitudes on childrearing practices, parental expectations for and satisfaction with day care services, and their degree of involvement in center activities. The findings from the Phase II Parent Interview are summarized below.

Most of the parents interviewed during Phase II had had some previous experience with day care, including center care, family day care homes and, less frequently, care by relatives inside and outside the child's home. Almost invariably they cited the need to work as their reason for using day care at all. A majority chose center care over other forms of care because of the superiority of its educational programs. In choosing a particular center, parents were influenced by the convenience of its location and hours of operation or relied on a friend's recommendation. Cost was a factor in the choice of center for only a few parents.

Parent Expectations

Parents stressed the need for trained and experienced caregivers, even as they expressed the belief that personal characteristics are more important determinants of caregiver quality than experience or education. In describing what they expected centers to do for their children, parents emphasized preparation for grade school, good supervision and discipline. Few parents expected to play an active role in the decisionmaking process at the center; less than half felt that it was important for parents to help hire staff.

Parent Involvement

Most parents were not actively involved in the activities of the day care center. They visited only to confer with center staff, to observe their children or to attend social events. One-fourth took advantage of the educational opportunities offered by the center through workshops, training sessions and parent education courses. Very few parents were employed at the center or played a major role in decisions concerning the center. Although many parents wanted more involvement, virtually none expressed interest in an expanded role in decisionmaking; rather, they voiced a desire to work as aides or to participate more actively in educational activities.

Parent Satisfaction

Although parents expressed general satisfaction with their centers, they identified several areas where improvements were needed. Some saw a need to strengthen and intensify the educational program of their centers. Others wanted caregivers with greater training or experience or additional space. Almost half expressed concern about the safety of the streets around the center; there was also concern about arrangements made by centers for the care of sick children.

Conclusions

In general, centers met parents' practical needs and provided care that parents deemed adequate or better.

Parent involvement in center activities and decision-making was not extensive and, while many parents desired increased involvement, few wished to participate further in making decisions. Two themes ran through many parents' responses to questions about their expectations, goals and evaluations of their day care arrangements: center care is seen as an educational opportunity, and staff characteristics are seen as important elements of quality care.

CHAPTER THREE: THE PHASE III SAMPLE

One hundred and twenty parents were contacted during Phase III; interviews were obtained with 112 of them. The interviews were distributed among twelve centers, four in each site, which received a portion or all of their funding through public subsidy. Thus, these parents, though they were a subsample of the Phase II interview sample, were not representative of the larger sample. Rather they represented a group whose views are especially relevant to federal day care policy, since their children were receiving care in subsidized centers.

All of those interviewed were female and were the mothers or grandmothers of children participating in the study. Two-thirds were black; most of the remainder were white, and a small number were Oriental or Native American. Approximately two-thirds had one parent or no parents in the home; the average size of families was 3.6. Almost two-thirds of the families had annual incomes of less than $6,000 (Figure 3.1). Almost two-thirds had considered other kinds of day care and other centers when they chose their centers; half chose their centers on the basis of a friend's recommendation.

The goals for this study were very different from the Phase II goals. The earlier parent interview sought to survey all parents of study children on a variety of issues. The Phase III interview focused on a set of topics that were most directly related to the regulation of day care by the federal government and, in particular, on parents' opinions about different policy variables. As the people most concerned about the impact of particular care arrangements on children, their judgments give some indication of the public acceptability of different regulations. While responses to all questions were precoded, a number of

Figure 3.1

DISTRIBUTION OF PHASE III INTERVIEW SAMPLE BY INCOME*

```
           |
         60|
P   o      T
e   f    50|
r          T          45
c   F    40|      |XXXXXXXX|
e   a      T      |XXXXXXXX|
n   m    30|      |XXXXXXXX|
t   i      T  21  |XXXXXXXX|
a   l    20|XXXXXXX|XXXXXXXX|   19
g   i      TXXXXXXX|XXXXXXXX|XXXXXXXX|
e   e    10|XXXXXXX|XXXXXXXX|XXXXXXXX|                                6
    s      TXXXXXXX|XXXXXXXX|XXXXXXXX|   4.5      4.5     |XXXXXXXX|
          0|XXXXXXX|XXXXXXXX|XXXXXXXX|XXXXXXXX|XXXXXXXXX|XXXXXXXX|
             0 -    $3,000-  $6,000-  $9,000-  $12,000-  $15,000
           $2,999   $5,999   $8,999   $11,999  $14,999      +

                                                        n=105
```

---

*For 10 families, no information on income was available.

questions were asked for additional verbatim responses. What is reported here are tabulated data and a selection of verbatim comments that add to our understanding of the precoded responses. In each case, the verbatim quotations represent the view of a number of parents rather than being an idiosyncratic response.

CHAPTER FOUR: PHASE III PARENT INTERVIEW FINDINGS

Parents' Goals for Children in Day Care

To look at the effects on children of variation in center characteristics, the NDCS selected a set of variables that reflected both the experience of the child in day care (process) and changes in the child as a consequence of that experience (outcomes). A number of considerations governed the choice of variables; perhaps the most important was that they should have legitimacy in the eyes not only of the policymakers and researchers but also of parents. The Phase III Interview offered parents an opportunity to tell us what they hoped the day care center would accomplish for their children. They were asked first to respond to an open-ended question on this subject. Table 4.1 shows the question and the responses. Three-quarters of the parents hoped that the center would teach their children school-related skills (Responses 8, 9 and 10). Over half cited a variety of social skills that they hoped the center would improve, such as cooperative or sharing behavior and independence. Responses included: "I want the center to expose her to a larger number of children, so she can learn to get along in a group." "The center will help her to grow up; she acted like a baby before." Clearly, the center, is perceived as a place where children learn to get along with a variety of people outside their immediate family.

A small number of parents saw the center as providing above all a safe and secure place to leave their child, so that they need not feel concern while working about the amount and quality of attention and supervision the child was receiving. A few wanted very specific help with the special problems of hyperactive, blind or non-English-speaking children.

Table 4.1

WHAT DID YOU HOPE THE CENTER WOULD DO FOR YOUR CHILD?

Responses	n	%*
1. Help my child to be less aggressive	3	3
2. Help my child to be more self-assertive/independent	18	16
3. Help my child to be more sociable	59	51
4. Help my child to share with other children	40	35
5. Help my child to be more obedient	8	7
6. Help my child to feel loved and secure	7	6
7. Teach my child self-help skills	16	14
8. Teach my child things he will need for school	64	56
9. Teach my child to develop language skills	18	16
10. Teach my child to be more interested in learning	6	5
11. Help my child's physical development	3	3
12. Help my child with special problems	8	7
13. Provide a safe environment	8	7

*Number and percentage of responses exceed 115 and 100 percent because parents were free to offer several responses.

Parents were then shown a list of things that the center might do for the child and asked to rank the three most important (Table 4.2). Preparation for school and development of a variety of social behaviors retained their importance for parents; their emphasis on school preparation and encouragement of independent, self-assertive behavior suggests that they see the center experience as helping their child take the first steps towards maturity. At the same time, a number of parents spoke of the center's role in making the child feel loved and secure.

General Day Care Preferences

To try to understand why parents chose center care rather than other forms of care and what aspects of it were important to them, we presented them with a series of forced-choice questions. Parents were asked to imagine themselves as advisors to a friend in need of child care and, in each case, to recommend one of two possible choices, giving the reasons for their choice. Their verbatim responses were later sorted into response categories.

The series of questions began with one that asked parents to recommend either family day care or center care to a friend. Since all were parents who had chosen center care after using other forms of care, it was not surprising that, for the most part, they recommended a day care center. Only six of the parents preferred family day care; one parent felt that it would depend entirely on the individual child, since not all three-year-olds are ready for a group experience. The largest number of those who recommended center care saw it as a more stimulating social experience for the child because of the number and variety of children s/he would meet (Table 4.3). Some saw it as "a preview of regular school" and felt that opportunities for school preparation were more limited in a family day home. There

214

Table 4.2

WHAT ARE THE THREE MOST IMPORTANT THINGS THE CENTER CAN DO FOR YOUR CHILD?

Responses	1	2	3
	n	n	n
1. Help my child to learn things he will need for school	22	19	23
2. Help my child to be more interested in learning	13	8	15
3. Help my child to develop language skills	6	13	5
4. Help my child to be more self-assertive/independent	16	21	17
5. Help my child to be more sociable	7	4	5
6. Help my child share with other children	10	13	9
7. Help my child to be more obedient	11	13	5
8. Teach my child self-help skills	8	11	12
9. Help my child to feel loved and secure	15	6	10
10. Help my child to be less aggressive	2	3	2
11. Help my child's physical development	2	2	6
12. Help with special problems	2	0	2
13. Not answered	1	2	4

n=115

Table 4.3

PRIMARY REASON FOR RECOMMENDING CENTER CARE

Reason	n
Staff professionalism and training	7
Emphasis on learning as opposed to play	20
More opportunities/variety of things to do	10
More interesting social experience	37
Centers are regulated	4
Dislike family day care	20
Don't know	10
	108*

---

*6 parents preferred family day care
1 parent gave a neutral response

was repeated emphasis in their comments on the "learning" that goes on in a center, as opposed to "play" in a family day care home. The structure of the center's day appealed to parents. Through many of their secondary responses ran the theme that the center is inspected, both by regulatory authorities and by the parents themselves if they wish: "There are guidelines for centers--you don't know what they eat or do in a home."

One group of parents expressed their preference for center care in terms of their dislike of family day care. These parents felt that they could not trust a family day care mother to care for their children unless the two families were friends. Otherwise: "the family day care mother would focus on her own children and not supervise or discipline the others." The small number of parents who preferred family day care felt that, on the contrary, children receive more individual attention in a home. One mother felt that "a three-year-old is too young to function in a group away from home. Day care centers don't have enough staff to pay attention to a three-year-old."

What should one look for in choosing a center? Most frequently, the answer was the philosophy of the center, its approach to child care and the kind of program it offered (Table 4.4). Other parents emphasized the cleanliness and adequacy of the physical environment, a focus on school preparation, or the qualifications and attitudes of the center staff. One group of parents talked more of strategies for finding out about the center: visiting; talking to other parents; and observing the child's interaction with other children. "Visit and observe; avoid too much structure or too much quiet." "Ask yourself, do the children enjoy it, are they happy and purposeful?" Only a few parents mentioned the cost of care; it should be remembered that most of these parents

Table 4.4

WHAT KINDS OF THINGS SHOULD YOUR FRIEND LOOK FOR IN CHOOSING A CENTER?

	n
Center philosophy and program	19
Emphasis on school preparation	10
Condition of facility, equipment, materials	14
Staff qualifications	9
Staff concern for children	9
Number of teachers/ratio of teachers to children	4
Center and class size	2
Convenience of center	6
Cost	2
Safety of the center	2
Availability and quality of meals	7
Other	13
No response	16
	115

received federally-subsidized care. While the same small fraction mention the safety of the center, it is likely that most parents take this for granted.

Preferences Concerning the Regulatory Variables

The NDCS staff examined a number of regulatable aspects of centers that seemed likely to affect children. Among these were the size of the group; the ratio of staff to children; and the training and experience of caregivers. The Phase III Parent Interview paired these variables in a number of ways and asked parents to choose the more important. We continued to use the device of asking the parents to imagine a friend in need of day care and to provide advice.

The first of these questions simply asked the parents to choose between a large group with an unexperienced teacher and a small group with a similar teacher and to give their reasons (the wording of the question is given in Table 4.5). Most parents chose the smaller group; those who chose the larger group felt that it exposed the child to a greater variety of children. Parents selected the small group feeling that it offered more opportunity for attention to individual children and for a closer relationship between teacher and children. Many parents pointed out that, lacking special skills, a teacher has an easier time with a smaller group. "A teacher can't handle a large group with no training." "The more children, the more unruly they can become and an inexperienced teacher will have more trouble."

Next parents were asked to choose between a large group led by an experienced teacher and a small group led by an inexperienced teacher; that is, they were asked whether they would trade their preferred group size for an experienced caregiver. Most, it seems, would do so; those parents

Table 4.5

SELECTED QUESTIONS FROM THE PHASE III INTERVIEW

Q. Suppose your friend has a choice between two centers. In one, her child would be in a large group, say 10-12 children, with a teacher with no special training or experience. In the other, the child would be in a small group, say 4-6 children, with a similar teacher. Which would you recommend that she choose?

Response	n
Large group	15
Small group	97
Don't know	3

n=115

Q. Suppose that your friend is choosing between two centers. Suppose the one with the larger group of children, 10-12 children, had a teacher with a great deal of day care experience and the smaller group of 4-6 children had a teacher with little or no day care experience. Which center would you recommend?

Response	n
Large group	88
Small group	26
Don't know	1

n=115

Q. The first center, with the group of 10-12 children, has a teacher with a master's degree in early childhood education. The second center, with the group of 4-6 children, has a teacher with a high school diploma. Both teachers have the same amount of experience. Which center would you recommend?

Response	n
Large group/special training	55
Small group/no special training	51
Neutral/don't know	9

n=115

Table 4.5 continued

Q. Suppose your friend has a choice between two groups that are the same size, ten children. One is led by a teacher just out of college with a degree in early childhood and the other has a teacher who has completed high school and has five years of day care experience. Which should she choose?

Response	n
Teacher with special training	13
Teacher with day care experience	93
Neutral/don't know	9

n=115

Q. Finally, suppose in one center twenty children are grouped together in one classroom with two teachers. In the other, the children are divided into two classes of ten, in two separate classrooms, with one teacher in each classroom. Which arrangement would you recommend?

Response	n
One large group	43
Two small groups	65
Neutral/don't know	8

who did choose the small group felt strongly that only in such a group could children receive individual attention. Most, however, disagreed: "If she is experienced, she would know how to work with kids as a group or individually." "Experience is important--she can do more with 10 or 12 kids than a less experienced teacher can with four or six kids." "Having experience, she would be able to make each child feel special and meet their needs." Though they understood that the larger group makes more demands on the teacher's organizational and management skills, they felt that an experienced teacher is better prepared to juggle the variety of demands with which she is faced.

When both teachers have the same amount of experience, would parents prefer a large group with a specially trained teacher or a small group whose teacher has no special training? Parents were almost evenly divided on this question; those who chose the larger group felt that an educated teacher is better able to handle special problems and apply teaching techniques: "She has a degree and it would help. She probably knows a little more about taking care of children." "A master's degree is better preparation and includes training in psychology." "Training would teach her how to deal with more children." Parents who chose the smaller group did so either because they continued to believe that it offered more opportunity for individual attention or because they did not believe that education, by itself, prepares a teacher to deal with small children. "Experience is the best teacher. A master's degree doesn't necessarily teach you how to relate to children." "Someone with a high school education may be wise with children." "Some people with degrees only have book learning." "Education doesn't guarantee common sense."

Parents were then asked to choose between a teacher with a newly acquired college degree and a teacher with a high school eduation and considerable experience, both working with groups of equal size. Three-quarters of the parents chose the teacher with substantial experience. In a variety of ways they expressed their conviction that the daily experience of dealing with children provides a teacher with as much knowledge as she could acquire by means of a college degree. "You can go through college and not be able to cope with children, whereas anyone sticking with day care five years has something going for her." "The school experience is different from the reality of a group of children." "Lack of scholastic training is made up for by on-the-job training. She can handle problems better."

Finally, parents were asked to choose between a center with a group of twenty children led by two teachers or a center in which twenty children were divided into two groups, in separate classrooms, with one teacher in each. Over one half preferred the smaller classes; they anticipated a good deal of confusion and noise or were concerned that the individual child would be lost and unheeded in a larger class. Slightly more than one-third of the parents chose the larger group, feeling that two teachers would work together more efficiently--"It's four eyes as opposed to two; if one doesn't see what is happening, the other will"--and would provide different ideas and perspectives--"It exposes the child to two viewpoints."

Parents' views about the importance of the different policy variables thus provide an important supplement to the NDCS effects findings. When asked to choose between two different staff/child ratios most parents preferred the high-ratio grouping because they felt that it offered more opportunity for attention to individuals and for a closer relationship between teacher and children. However, it was apparent from parents' responses that they were not considering the classroom ratio in isolation, but rather its interaction with another independent variable, teacher experience.

An inexperienced teacher, they stressed, is unable to handle the larger number of children without sacrificing important aspects of care. Their choice of the higher-ratio grouping was a safeguard against inexperienced teachers.

This reasoning became even clearer when parents were asked to choose between a group with their preferred ratio supervised by an inexperienced teacher, and a lower-ratio grouping with an experienced teacher. Most chose the low-ratio grouping because they felt that a teacher who has had substantial day care experience is able to organize a larger group of children so that the quality of care is the same as or better than that provided by an inexperienced teacher with a small number of children. Parents were less impressed with the benefits of formal education and were unwilling to exchange a teacher with day care experience for one with a college degree. It should be stressed that these parents understood "education" to mean formal classroom courses in, for example, child development; many parents' comments in the interviews suggested that the kind of practical classroom apprenticeship that is a part of many specialized training programs would count as "experience" in their eyes.

Parents' Knowledge of the Child's Day

In Phase III parents were asked to talk a little about their child's day at the center, to say what they felt the child was learning, what they liked best and liked least when they observed. Although parents were reasonably familiar with some of the events of the child's day, their views on what the child was learning were derived primarily from conversation with or observation of the child at home. Occasionally a parent would explain that they did not have a strong sense of this because their child was very quiet and talked very little about the center. Others were pleased by the child's expanding vocabulary, improved pronunciation and increase in language skills. Parents' comments suggest that the effects of art and music sessions at the center are observed at home; children bring home their drawings and the songs and rhymes they have learned. Parents see, and credit the center for, increasing social maturity and a growing ability to deal with other children and adults (Table 4.6).

Less than two-thirds of the parents had observed at the centers and were thus able to discuss their feelings about what they had seen. Most of those who observed were impressed by the way teachers organize the group so that at any one time most children were playing happily while one or two received special attention.

"The teacher sits with the children and gives them a lot of individual attention." "I liked seeing Sam involved with other children; he was so interested in what they were doing and participated so well." Others liked the teacher's ability to control the children and maintain order while remaining warm and patient. "I liked seeing him sit and listen so attentively to the teacher." "I liked the way the teacher disciplined the kids; she tried to find out why they were fighting." "The teacher had a lot of patience--paid attention to each child."

For a third group of parents, the experience of observing taught them new and pleasant things about their own child. "I liked to see how involved he gets in the things he does." "Her friends really like her and got excited when she came in; they value her opinions." "I was pleased to see how contented she was--it eases my mind when I go to work."

Table 4.6

WHAT DO YOU THINK YOUR CHILD IS LEARNING AT THE CENTER?

Response	n
Language skills	32
Math skills	11
Abstract information	8
Expressive skills	25
To deal with others	19
To be more considerate	1
To be more assertive/outspoken	2
To be more obedient	
Other social skills	7
Motor skills	3
Other	1
Unresponsive/don't know	6

n=115

Less than half the parents had seen things that troubled them while they observed at the center. The confusion at the end of the day, when clothes are misplaced and children are more likely to fight, concerned them. The deficiencies of the physical setting were often mentioned--the stuffy air and crowded atmosphere of the napping room, the need for more toys and equipment. Ringworm, colds and infections in other children were also a source of concern.

Parent-Center Communications

More than half of the parents talked with their child's teacher at least once a week (Table 4.7); contact with the center director was somewhat less frequent. For the most part, parents felt that the teacher was available when they needed to talk to her and that they did not need more time with her.

Meetings with teachers most frequently took place when parents picked up their child from the center, or, less often, after they brought the child to the center; only infrequently were the meetings by special appointment. Conversations with teachers focused on the child's overall progress or on specific areas in which there were problems to be worked on or progress to be reported (Table 4.8). Parents' comments suggest that teachers make an effort to balance a discussion of problems with a recognition of progress. "We talk about problem areas that need improvement, but she mentions pleasing areas also." "His teacher tells me what progress he is making towards the goals she has in mind." "We talk about what help he needs from the home, things I can do to help him improve."

A majority of parents said that the advice they got from the teacher helped them in dealing with their child at home. In general, the advice was of two kinds: how to

Table 4.7

HOW OFTEN DO YOU TALK WITH YOUR CHILD'S TEACHER? WITH THE CENTER DIRECTOR?

Response	With Teacher	With Director
Daily	30	21
About twice a week	20	15
Weekly	13	12
Once or twice a month	25	33
Rarely	26	33
Never	1	1

n=115

Table 4.8

WHAT KINDS OF THINGS DO YOU MOST OFTEN TALK ABOUT WITH THE CHILD'S TEACHER?

Response	n
Child's overall progress	43
Cognitive problems/progress	20
Social problems/progress	19
Other problems	7
Center activities	12
Other topics	12
Non-responsive	1

n=115

reinforce at home what the child was learning at school; and how to deal with behavior and discipline problems. While some teachers suggested learning activities for the child that closely resembled center activities (such as cutting and pasting paper shapes), others recognized the unique aspects of the home and suggested to parents that they talk to children about the smell and taste of food, about the function of household objects and how they work. Teachers' suggestions about behavior problems included ways of dealing with tantrums and bedwetting as well as the more serious problems of hyperactivity and physical handicaps. Often, conversations with the teacher dealt with what was reasonable to expect of the child, what normal developmental patterns are and the importance of praise and encouragement. In a few cases, teachers gave parents home phone numbers so that they could call at any time for help with problems. Only a few parents felt that there was information they needed that the teacher did not provide, sometimes because she was not able to. "I'd like to know how to help him learn to read and they don't have the training to help me do that."

Parents were divided on the question of consistent discipline at home and at the center. Slightly more than half felt very strongly that the discipline should be consistent. "I am a working mother and I can't do it alone. I want the center to reinforce what I say so that it doesn't break the pattern." "He is used to certain kinds of discipline at home and I want that to continue at the center." Those who felt that the discipline should not or need not be consistent were, for the most part, expressing their sense that the teacher might have more effective methods of discipline. "She is fair, and can do better than I do." "We are two different people and have different but effective ways. I spank him and she doesn't because she has other means." "Children have to deal with all kinds of people and they need to know that there are other ways to deal with things." Some parents pointed out that although they spanked their children, they did not want them spanked by other people.

Few parents disagreed with teachers on how to care for their children. Those who did had unique needs that they felt were not being met. "I'd like the teacher to read stories about happy single-parent families." "She should tell the children that blacks and whites do associate with each other instead of remaining isolated." Often these parents were hesitant to talk with the teacher about their differences, usually because they feared outright rejection of their ideas.

Most parents agreed that, because the teacher has a large group to care for, her ways of managing children must be somewhat different. However, only a few felt that this necessitated reduced attention to the child's individual needs.

Parent Involvement

During Phase II, parents were asked about the ways in which they were involved in the center and the ways in which they would like to be involved. Only a small fraction of those interviewed had or wanted any involvement in the decisionmaking processes of the center. Their visits to the center were mainly to talk to staff members and to observe their child; a large number wanted to spend more time observing.

The Phase III sample was, in this respect, a more active and politically aware group. Ten percent sat on advisory boards and helped make decisions about the center program and the hiring of staff. Half of those interviewed said that their centers had advisory boards, and felt that parent participation on these boards was important. In general, they felt it was more important for parents to help make hiring and program decisions than to review the budget; perhaps this task seems to them to require specialized financial skills. More than a third of the parents did not know whether their centers had advisory boards; these same parents had few strong feelings about parent participation on such boards. Once again, parents usually visit the center to talk to center staff or to observe, and centers clearly encourage this; however, almost half attended parent educational meetings and about a quarter spent time as a volunteer in the classroom or helping out on field trips. Very few were employed as paid aides in the center. Parents seemed satisfied, on the whole, with the extent of their involvement, although a number of them would like to be more involved in classroom activities. Many parents pointed out that, since they are working, their ability to participate is limited. If centers hold board meetings or parent-teacher conferences during the day, only a few parents can take part in these activities. Those parents who expressed a desire to volunteer in the classroom were realistically aware that working limited their opportunities for such involvement.

Parents' Use of Center Services

Parents identified an array of services provided by centers. While less than half said that their center provided legal or job-related help for families, a majority said that the center offered health services. Counseling, social services and transportation to the center were frequently mentioned. Other services were more informal and often unique to a particular center. Some centers provided holiday food for needy families, in addition to helping them obtain food stamps. Another center provided financial counseling and helped parents with their tax forms. Occasionally, centers provided emergency shelter for a family or kept children over a weekend in an emergency. Most frequently, parents were told of available services by the teacher or other center staff. Sometimes the center's social worker provided all new parents with a list of center services and resources. Newsletters and more informal communications on the bulletin board kept parents aware of new or existing services. Less than half of the parents interviewed had used a service offered by the center; most often this was a health or dental service. Asked what additional services, if any, they would like the center to offer, half felt that enough services were available; the remainder were concerned that health and social services be strengthened to meet some emerging needs--to deal with child abuse and to help single parents with childrearing problems. Several parents wanted some additional parent education services focused on specific problems, such as common childhood illnesses.

Parent responses make it clear that communities in each of the three cities possess a variety of resources to help them. Local health clinics and legal aid services were most frequently mentioned, but emergency food and clothing resources, family counseling centers, and protective service agencies were also identified.

Conclusions

The most striking finding of the Phase III Parent Interview supports the Phase II findings: parents saw center day care as an experience that prepares the child cognitively and socially for school. Their goals for children echoed the philosophies articulated by many center directors. Parents stressed center philosophy, staff experience and availability of dental and health care as the center charactersitics that most contribute to the desirability of a particular day care center. Small groups and high staff/child ratios were also seen as desirable, although parents generally felt that experienced staff could handle larger groups and lower ratios successfully. Although a number were active in the planning and administration of their centers, most parents wanted more involvement in the educational activities in which their child participates, perhaps to understand the learning process better, so that they could more effectively help the child at home. The communication between this group of parents and their children's teachers seemed to be generally satisfying; in some instances, there existed a warm and trusting relationship that allowed both parents and teachers to learn from and depend upon each other.

The Classroom Environment Study

Barbara Dillon Goodson
with
Lorelei R. Brush
Catharine Barclay

TABLE OF CONTENTS

Page

CHAPTER ONE: INTRODUCTION 219

CHAPTER TWO: CLASSROOM STRUCTURE OF DAY CARE CENTERS 220
  Number of Classrooms 221
  Ages of Children Served 223
  Changes in Classroom Structure 224
  Profiles of Centers by Classroom Structure 225

CHAPTER THREE: CLASSROOM COMPOSITION 227
  Classroom Staff 227
    Staff Size 227
    Full-Time and Part-Time Staff 227
    Lead Teachers and Aides 228
    Aide Responsibility 230
    Staff Stability 232
  Children 237
    Center Enrollment 237
    Classroom Enrollment 237
    Full-Time and Part-Time Children 240
    Race and Sex of Children 241
  Staff and Children 242

CHAPTER FOUR: CHARACTERISTICS OF DAY CARE CENTER DIRECTORS 245

CHAPTER FIVE: CENTER PHILOSOPHY 248
  Director's Goals 248
    Orientation toward Individual Children 248
    Group Orientation 248
    Cognitive Emphasis 248
    Ethnic/Religious Emphasis 249
    Guidance Emphasis 249
    Protective Emphasis 249
    Emphasis on Behavioral Supervision 249
  Classroom Goals 251

CHAPTER SIX: MEASURES OF SPACE AND THE PHYSICAL ENVIRONMENT 253
  Center Space Plans 253
  The Physical Environment Inventory 254
  Factors from the CDA Checklist 257
  Creating an Index of the Physical Environment 258
  Summary 260

CHAPTER ONE: INTRODUCTION

The National Day Care Study examined the links between regulatable aspects of day care centers and outcomes for children. While a major goal of the study was to determine how, if at all, staff/child ratio, group size and caregiver qualifications influenced the development of children in day care centers, the study staff also examined such things as the classroom staffing structure, directors' characteristics, centers' physical environment, and center philosophy. Although these aspects of the day care environment were examined primarily to determine their impact on child outcomes, a secondary result was the richness this examination brought to our understanding of the day care environment. This paper is essentially a description of the classroom environment in the NDCS centers.

The paper itself covers four major topics. The bulk of this paper is devoted to an analysis of classroom structure within centers (Chapter 2), staffing structure within the classroom (Chapter 3), and, to a somewhat lesser extent, to a discussion of the center director's characteristics (Chapter 4). The results of an investigation of center philosophy are reported (Chapter 5), as are the findings concerning the physical space available to children within a given center (Chapter 6).

CHAPTER TWO: CLASSROOM STRUCTURE OF DAY CARE CENTERS

The analysis of the classroom structure was organized around the independent variables used in the analyses of classroom process (i.e., number of staff and children). This analysis was intended to provide a picture of the natural variation in classroom, child and staff characteristics, and of interrelations among the descriptive variables. The classroom environment study covers three major topics:

- classroom structure of the 57 centers;
- classroom composition (staff and child characteristics); and
- classroom program and orientation.

Classrooms were considered at two time points--October 1976 and April 1977. Most descriptions are based on April data. In addition, changes from October to April in both classroom structure of the centers and composition of the classrooms were examined to determine whether the degree of discontinuity in classrooms is related to the day care process and its outcomes. Data on classroom composition were taken from rosters of staff and children completed in October 1976 and April 1977. Roster information included background characteristics and schedules for individual staff members and children. Questionnaires completed by all lead teachers in target classrooms provided the information on classroom organization and programs.

Centers were categorized two ways: by auspice--whether they were privately operated or sponsored by a public agency, and by whether or not they enrolled any federally subsidized children. With only a few exceptions, most of the centers serving federally subsidized children were sponsored by a community agency and were classified as public centers. Thus classroom characteristics linked to public auspice were almost always also linked to the presence of federally funded children. In a few cases, however, a distinction could be made between public or private auspice regardless of the presence or absence of a federally funded enrollment or between centers serving federally funded children and those serving only privately paying children regardless of auspice. Thus findings about classroom patterns are reported both by center auspice and by types of center enrollment.

Number of Classrooms

There were approximately 250 classrooms in the 57 study centers during Phase III--242 in October and 246 in April. Table 1 presents the distribution of these classrooms and centers across the sites, and Table 2 shows a detailed breakdown of the number of classrooms in centers, by site. For all sites, the mean number of classrooms per center was 4.24 in October and 4.31 in April; the modal structure was the five-classroom center in October and the four-classroom center in April.

In general, Atlanta centers had the largest number of classrooms, and Detroit centers the smallest. In October, for example, the mean number of classrooms per center in Atlanta was nearly 5, compared to 3.6 in Seattle and only 2.9 in Detroit. Modal center sizes show a similar pattern. The most frequent center structure in Atlanta was the four-classroom center, and 75 percent of Atlanta centers had four to six classrooms. In Seattle, the modal center size was three classrooms, and more than half of all centers had three classrooms or less. Detroit centers had the most uneven distribution of center structures and the largest number of small centers. The most common center structure in Detroit was the one-classroom center, and more than half of Detroit centers had one or two classrooms.

Table 1

NUMBER OF CENTERS AND CLASSROOMS BY SITE, OCTOBER AND APRIL

October	Atlanta	Detroit	Seattle	All Sites
Number of Centers	28	13	16	57
Number of Classrooms	139	45	58	242
Mean Number of Classrooms per Center	4.95	2.92	3.62	4.24
Modal Number of Classrooms per Center	4	1	3	5
April				
Number of Centers	28	13	16	57
Number of Classes	138	49	59	246
Mean Number of Classrooms per Center	4.92	3.23	3.68	4.31
Modal Number of Classrooms per Center	4	1	3	4

Table 2

CLASSROOMS PER CENTER BY SITE, OCTOBER AND APRIL

	Number of Classrooms per Center								
October	1	2	3	4	5	6	7	8	9
No. of Atlanta Centers	-	1	3	8	7	6	-	2	1
No. of Detroit Centers	5	3	1	-	2	-	2	-	-
No. of Seattle Centers	3	-	6	2	3	1	-	1	-
No. of All Sites	8	4	10	10	12	7	2	3	1
April									
No. of Atlanta Centers	-	1	2	9	7	7	-	1	1
No. of Detroit Centers	4	3	2	-	2	-	1	-	1
No. of Seattle Centers	3	1	4	3	2	2	-	1	-
No. of All Sites	7	5	8	12	11	9	1	2	2

Ages of Children Served

The following types of centers were identified among the 57 study centers:

- centers with preschool classrooms only (mean age range of 2.5 - 4.9 years);
- centers with preschool and school-aged classrooms (mean age range of 2.5 - 6.0+ years);
- centers with preschool and infant classrooms (mean age of 0.1 - 4.9 years); and
- centers with infants, preschool and school-aged classrooms (mean age range of 0.1 - 6.0+ years).

The most common age distribution in the 57 centers was 2.5 - 6.0+ years for nearly all sites in both October and April (Table 3); about half of all centers were composed of preschool and school-aged classrooms. Atlanta was the only site with a substantial number of centers that served infants, and relatively fewer centers there served only preschool children.

Table 3

CENTER AGE RANGES BY SITE, OCTOBER AND APRIL

October	Preschool Only (2.5 - 4.9 yrs.)	Preschool & School Age (2.5-6.0+yrs.)	Preschool & Infant (0.1 - 4.9 yrs)	Infant, Preschool & School-Aged (0.1-6.0 yrs)
No. Atlanta Ctrs.	4	7	4	10
No. Detroit Ctrs.	6	6	1	-
No. Seattle Ctrs.	3	13	-	1
All Sites	13	25	8	11
April				
No. Atlanta Ctrs.	3	11	7	7
No. Detroit Ctrs.	5	6	1	1
No. Seattle Ctrs.	3	12	-	1
All Sites	11	29	8	9

Changes in Classroom Structure

From October to April changes were made in classroom structure in 13 of the study centers. These changes fall into five types:

- established classes underwent substantial changes in staff or enrollment;
- new classes were formed;
- existing classes were dropped;
- two existing classes merged to form one class; and
- an existing class split to form two classes.

The most frequent change was the reorganization of classes as a result of substantial shifts in staff and/or children. Such reorganization occurred in 17 classrooms; 50 percent of these changes took place in three- and four-year-old classrooms. Six entirely new classes were formed during the year; the majority of these were for school-aged children. Two new classes of three- and four-year-olds were formed when two existing classrooms split. Five classes were dropped during the year, the majority of which were infant and toddler classes. Only one class was "lost" when two existing classrooms were merged. Half of the changes in structure took place in December, after the centers had been operating long enough to assess enrollment demands for the year and reorganize classroom structure accordingly.

Seattle centers were the least stable; there were structural changes in 31 percent of Seattle centers, while changes occurred in less than 20 percent of Detroit and Atlanta centers. Centers with changes in classroom structure tended to be larger than average, with a mean enrollment of 71 children. Likelihood of change was also related to the auspice and funding source of the center. Overall, more private centers experienced changes than did public centers (38% vs. 17%) and more non-federally funded centers experienced changes than did federally funded centers (28% vs. 18%). Private centers not serving federally subsidized children had proportionally the greatest number of changes in classroom structure (42%).

Profiles of Centers by Classroom Structure

In all three sites, centers could be categorized as being typically small, medium or large. The typical small center had one or two classrooms with a total enrollment of less than 40 children. The typical medium-sized center had three or four classrooms and a total enrollment between 40 and 65 children. The typical large center had five or more classrooms and 60 or more children.

Among the study centers, a typical small center would most probably be found in Detroit. The center would serve only preschool children. The center would have an enrollment of 38, and the mean age in the center would be 4.5 years. There would be very little structural change in such a center over the year.

The typical medium-sized center would be found most frequently in Seattle, with three classrooms and an enrollment of 59 children. There would be two preschool classrooms--a three-year-old classroom of 18 children and a four-year-old classroom of 25 children. An older class of school-age children would serve 16 five-year-olds. The center would not serve infants. The classroom structure of this center would be unlikely to change during the year.

The typical large center would probably be located in Atlanta, with an enrollment of around 80 children in five classrooms. One classroom would serve 14 children under three years of age--mostly toddlers or infants. Three classrooms would serve three- and four-year-old preschoolers: one classroom would serve 17 three-year-olds and the other two classes would serve primarily four-year-olds and young five-year-olds. The enrollments in these classrooms would be 15-18 children. The fifth classroom would serve 18 five-year-olds. It is likely that this large center would undergo structural changes over the course of the year, including changing staff or shifting children in existing classes throughout the year and adding a new class in the spring.

CHAPTER THREE: CLASSROOM COMPOSITION

Classroom Staff

Information on classroom staff was collected at five points in Phase III. At each point staff members were identified with a classroom (or classrooms), classified as full- or part-time, and as having lead teacher or aide responsibilities in the classroom. The following discussion is focused on staff in the centers in April 1977. Most of the discussion concerns target staff only--staff in classrooms with three- and four-year-olds. The description of staff includes:

- total number of staff assigned to classrooms;
- numbers of full- and part-time staff;
- numbers of lead teachers and aides; and
- aide responsibility.

Staff Size

There were approximately 650 staff in the 57 study centers; about four hundred of these staff were in the 137 target classrooms (Table 4). Across all sites, the mean number of caregivers in target classrooms was 2.9; the two-caregiver classroom was the most common configuration in all three sites. The three-caregiver classroom was second most common in Atlanta and Seattle, whereas single-caregiver classrooms were relatively uncommon (Table 5). In Detroit, however, the one-caregiver classroom was second in frequency, and there was also a slightly higher proportion of classrooms in Detroit with five or more caregivers. The two Detroit classrooms with eleven assigned caregivers, however, were single-classroom centers.

Table 4

NUMBER OF STAFF ASSIGNED TO TARGET CLASSROOMS, APRIL

	Atlanta	Detroit	Seattle	All Sites
Number of Classes	72	31	34	137
Number of Staff	203	96	100	399
Mean Number of Staff/Class	2.8	3.1	2.9	2.9
Modal Number of Staff/Class	2	2	2	2

Full-Time and Part-Time Staff

In each of the centers, there were both full- and part-time class assignments. For the NDCS, full-time was defined as more than 30 hours per week in class, and part-time was less than 30 hours per week. Across all sites, 58 percent of the target staff were full-time. Atlanta had a high proportion of full-time staff (70%), whereas only 43 percent of the staff in Seattle and Detroit classrooms were full-time (Table 5). In all three sites, the majority of the classrooms had at least one full-time staff member (Table 6). The modal number of full-time staff per classroom was two in Atlanta and Seattle and only one in Detroit. Detroit also had the highest proportion of classrooms with at least one part-time staff member--87 percent, compared with 82 percent of Seattle classrooms and 53 percent of Atlanta classrooms--and the highest proportion of classrooms with three or more part-time staff (Table 6).

Table 5

FULL-TIME STAFF IN TARGET CLASSROOMS, APRIL

	Atlanta	Detroit	Seattle	All Sites
Percent Full-Time Staff	70%	43%	43%	58%
Mean Number Full-Time Staff per Class	2.0	1.2	1.2	1.4
Modal Number Full time Staff per Class	2	1	2	1

The proportion of full-time staff was related to center auspices and funding sources. Public centers and

Table 6

NUMBER OF STAFF ASSIGNED TO TARGET CLASSROOMS, APRIL

			Number of Full-Time Staff								
	0	>1	1	2	3	4	5	6	Total Class-rooms	Mean # FT Staff	Modal # FT Staff
No. Atlanta Classes	5	67	23	24	11	6	1	2	72	2.0	2
No. Detroit Classes	8	23	17	3	0	0	2	1	31	1.2	1
No. Seattle Classes	11	23	7	12	4	0	0	0	34	1.2	2
All Sites	24	113	47	39	15	6	3	3	137	1.4	1

			Number of Part-Time Staff						
	0	>1	1	2	3	4	Total Class-rooms	Mean # PT Staff	Modal # PT Staff
No. Atlanta Classes	34	38	22	12	4	0	72	.8	0
No. Detroit Classes	4	27	17	3	2	5	31	1.6	1
No. Seattle Classes	6	28	11	12	1	4	34	1.3	2
All Sites	44	93	50	27	7	9	137	1.4	1

centers with federally funded children tended to have more full-time staff. Since these types of centers were most common in Atlanta, this may account for the higher proportion of full-time staff in Atlanta.

Lead Teachers and Aides

Teachers and aides were the two major types of classroom staff. For the most part, teachers had the primary responsibility for classroom activities, and aides provided assistance to the teachers. Eighty-five percent of the 137 target classrooms had at least one aide, and 44 percent had more than one aide (Table 7).

In all three sites, the most common staffing arrangement was one teacher and one aide (Table 7). This configuration occurred in 30 percent of all classrooms. Although it was the most typical staff configuration in each of the sites, the configuration appeared in a lower proportion of Detroit classrooms (22%) than in Atlanta or Seattle classrooms (31% and 35%, respectively). In Detroit and Seattle, the second most common classroom configuration was multiple lead teachers, with or without aides. In Atlanta, the second most common configuration was the classroom with one teacher and multiple aides. In general, an increased number of staff in a classroom meant the addition of aides rather than multiple lead teachers.

Lead teachers were significantly more likely than aides to work full-time in a classroom. Across all sites, 68 percent of the lead teachers worked in classrooms full-time, compared with 36 percent of the aides. This teacher/aide difference was true for all three sites.

Aide Responsibility

When lead teachers in the target classrooms were questioned about the roles aides played in the classroom in

Table 7

STAFF ARRANGEMENT IN APRIL TARGET CLASSES

	Aides Only	Teachers w/out Aides			Teachers with Aides			Teachers with Multiple Aides				
No. Teachers	0	1	2	3	1	2	3	1	1	1	2	
No. Aides	≥ 0	0	0	0	1	1	1	2	3	≥ 4	≥ 2	ALL
No. Atlanta Classes	1	10	3	3	22	4	3	15	5	4	3	72
No. Detroit Classes	0	7	5	1	7	3	2	1	1	0	4	31
No. Seattle Classes	1	3	3	0	12	2	1	4	4	2	2	34
No. All Sites	2	20	11	3	41	9	6	20	10	6	9	137
Proportion All Sites	.01	.26			.41			.44				

planning and carrying out structured activities, aides were described as assisting in planning the creative and group learning activities in less than half of the 100 target classrooms with aides. In Detroit nearly 90 percent of the lead teachers reported that they alone planned group learning and creative activities. In Atlanta and Seattle as well, the majority of the lead teachers reported doing all planning of these activities themselves. Across all sites, directors rarely participated (1% of classrooms). In general, aides in classrooms with more than one aide were more active participants in both planning and carrying out group activities. The three-caregiver classroom, with one teacher and two aides reported the highest level of participation by aides. It was reported that, typically, aides in two-caregiver classrooms did not assist in planning the structured activities, although they often directly assisted in carrying out these activities.

The level of aide participation was related to center auspice and funding. Aides in public centers more often had active roles than aides in private centers, and aides in federally funded centers were often more active than aides in non-federally funded centers.

Staff Stability

The issue of staff stability in day care classrooms is of particular interest. Some state regulations recommend continuity of caregivers to make day care more similar to a stable home environment. Some psychologists have hypothesized that the presence of a stable caregiver to whom the child can form an attachment is a crucial variable in the effect of day care. The stability of caregivers in NDCS classrooms was investigated through staff rosters for October 1976 and April 1977. The proportion of staff who remained in the same classroom over this period ("stable staff") was computed for each classroom. Stability was computed both for all staff and for teachers and aides separately. Non-stable staff were assigned as moving from class to class within one center or moving from one center to another.

Nearly half of the target classrooms had no changes in staff from October to April (Table 8). However, eight percent of the classrooms had a complete turnover in staff, and in 28 percent of the classrooms at least half of the staff changed during the year. The highest proportion of classes with 100 percent stable staff was found in Atlanta--57 percent of the target classes in Atlanta had no staff changes. In both Detroit and Seattle approximately 40 percent of the target classrooms were completely stable.

Table 8

PROPORTION OF STABLE STAFF, APRIL TARGET CLASSROOMS

	Proportion Stable Staff						
	0%	1-25%	26-50%	51-75%	76-99%	100%	TOTAL
No. Atlanta Classes	2	0	16	12	0	39(.57)	69
No. Detroit Classes	4	0	8	3	2	11(.39)	28
No. Seattle Classes	4	1	8	5	0	10(.38)	28
No. All Sites	10(.08)	1(.02)	32(.26)	20(.16)	2(.01)	60(.48)	125

There was more stability among lead teachers than among aides. Nearly 75 percent of the target classes had the same lead teacher(s) from October to April (Table 9); only 44 percent had the same aide(s) (Table 10). In each of the three sites, over half of the target classrooms had the same lead teacher(s) throughout the year, and only twenty percent of the target classrooms had complete turnover in lead teachers. Atlanta had the highest number of classrooms with no turnover, for both lead teachers and aides. In

Table 9

PROPORTION OF STABLE APRIL TEACHERS, APRIL TARGET CLASSROOMS

	0%	1-25%	26-50%	51-75%	76-99%	100%	TOTAL
No. Atlanta Classes	10(.14)	0	1	0	0	58(.84)	69
No. Detroit Classes	9(.32)	0	3	1	0	15(.54)	28
No. Seattle Classes	7(.26)	0	0	1	0	19(.68)	27
No. All Sites	26(.21)	0	4	2	0	92(.74)	124

Table 10

PROPORTION OF STABLE AIDES, APRIL TARGET CLASSROOMS

	0%	1-25%	26-50%	51-75%	76-99%	100%	TOTAL
No. Atlanta Classes	17	1	6	3	0	31(.53)	69
No. Detroit Classes	10	0	2	1	0	6(.32)	19
No. Seattle Classes	12	0	3	3	0	9(.33)	27
No. All Sites	39(.38)	1	11	7	0	46(.44)	104

addition to staff turnover, some classrooms increased or decreased in number of staff. Here again lead teachers were more stable than aides. Twenty-five percent of classrooms had changes in the number of assigned lead teachers, whereas about half of the classrooms had changes in the number of aides. Overall, the stability of lead teachers was not significantly related to the stability of aides (r = .12); that is, classrooms with high turnover in lead teacher(s) did not necessarily also have high turnover among aides.

The majority of changes in classroom staff involved moves from one center to another. Only about 30 percent of the staff changes were within-center transfers; lead teacher transfers were more often within-center than aide transfers (40% versus 32%).

Staff stability with teachers and aides combined was strongly related to center auspices and funding source (Table 11). Classrooms in public centers had significantly higher staff stability than those in private centers (74% versus 64%), and classrooms in federally funded centers had more stable staff than classes in non-federally funded centers (78% versus 65%). Classes in public, federally funded centers demonstrated the highest level of staff stability, and those in private, non-federally funded centers the lowest. The federally funded centers had higher stability than the non-federally funded centers for both public and private auspice.

When lead teachers were considered alone, stability was related only to auspice with public centers having more stable lead teachers. Federally funded centers had higher, but not significantly higher, stability than non-federally funded centers (Table 12). For aides, the classrooms in federally funded centers had significantly higher average aide stability than did classrooms in non-federally funded centers, while the auspice of the center did not have a significant effect (Table 13). Only the classrooms in public, federally funded centers had relatively stable aides.

Staff stability was not strongly related to either staff size or the number of children in the class. However, stability was affected by the staffing arrangement in a classroom. Lead teachers in classrooms with aides were more stable, on the average, than lead teachers in classrooms without aides (Table 14). In classes without aides, the lead teacher in a single-caregiver classroom was

Table 11

MEAN PROPORTION OF STABLE STAFF IN APRIL TARGET CLASSROOMS AS A FUNCTION OF CENTER CHARACTERISTICS

		Federally Funded	Non-Federally Funded	
Auspice	Public	.77 (n=62)	.67 (n=37)	.74
	Private	.82 (n=6)	.60 (n=26)	.64
		.78	.65	.71

Table 12

PROPORTION OF STABLE TEACHERS IN APRIL TARGET CLASSROOMS AS A FUNCTION OF CENTER CHARACTERISTICS

		Federally Funded	Non-Federally Funded	
Auspice	Public	.84	.78	.82
	Private	.60	.62	.62
		.82	.72	.77

Table 13

PROPORTION OF STABLE AIDES IN APRIL TARGET CLASSROOMS AS A FUNCTION OF CENTER CHARACTERISTICS

(n=104)

		Federally Funded	Non-Federally Funded	
Auspice	Public	.65	.40	.56
	Private	.43	44	.43
		.64	42	

more likely to have remained in her classroom over the year than were lead teachers in classrooms with two or three lead teachers. In classrooms with aides, the average stability of the aides was higher in classrooms with more than one aide.

Children

Information on children enrolled in each center was collected at five time points during Phase III. Based on this information, the following discussion of children covers these topics:

- center enrollment;
- classroom enrollment;
- number of full- and part-time children in classrooms; and
- race and sex of children.

Center Enrollment

In April 1977 there were approximately 4,500 children enrolled in the 57 study centers (Table 15), about 3,100 of whom were target children. Fifty-three percent of the children were enrolled in Atlanta centers, 20 percent in Detroit and 27 percent in Seattle. Across all 57 centers, the average enrollment was 80 children per center. Average center enrollment was highest in Atlanta, at 87.2 children, and lowest in Detroit, at 69.7 children. The center with the greatest total enrollment was located in Seattle and served 203 children; the smallest center enrollment was located in Detroit and served 34 children.

Classroom Enrollment

Classrooms in the study centers were placed in one of four categories according to class size:

- enrollment of less than 10 children;
- enrollment of 10 to 20 children;

Table 14

PROPORTION OF STABLE STAFF AS A FUNCTION OF STAFF ARRANGEMENT, APRIL TARGET CLASSROOMS

N of Teachers	0	1	1	1	1	1	2	2	2	3	3	3
N of Aides	≥ 1	0	1	2	3	≥ 4	0	1	≥ 2	0	1	≥ 2
N Classrooms	1	13	44	20	10	6	5	11	11	2	2	1
Average Teacher Stability	-	.69	.77	.83	.90	.80	.20	.86	.82	.50	1	.67
Average Aide Stability	.50	-	.41	.71	.63	.60	-	.36	.59	-	1	.33

Table 15

ENROLLMENT BY SITE, APRIL

	Atlanta	Detroit	Seattle	All Sites
Total Enrollment	2,442	906	1,225	4,573
Percent of Total Study Enrollment	53.4%	19.8%	26.8%	100.0%
Mean Enrollment Per Center	87.2	69.7	76.5	80.2

- enrollment of 21 to 30 children; or
- enrollment of more than 30 children.

In October and April, the majority (about 60%) of the classrooms enrolled between 10 and 20 children (Table 16). Classes of more than 30 children were least common, in general accounting for less than 10 percent of the classrooms. Very small classrooms (fewer than 10 children) and classrooms of 21 to 30 children were not uncommon.

Table 16

NUMBER OF CHILDREN ENROLLED PER CLASS, OCTOBER AND APRIL

	Less than 10 Children	10 - 20 Children	21 - 30 Children	31+ Children	Total
October					
No. Atlanta Classes (%)	26 (.19)	81 (.59)	22 (.16)	8 (.06)	137
No. Detroit Classes (%)	7 (.16)	27 (.60)	7 (.16)	4 (.09)	45
No. Seattle Classes (%)	7 (.13)	33 (.59)	10 (.18)	6 (.11)	56
No. All Site	40 (.17)	141 (.59)	39 (.16)	18 (.08)	238
April					
No. Atlanta Classes (%)	2 (.18)	83 (.61)	22 (.16)	6 (.04)	136
No. Detroit Classes (%)	3 (.06)	31 (.63)	11 (.22)	4 (.08)	49
No. Seattle Classes (%)	6 (.19)	33 (.58)	11 (.19)	7 (.12)	57
No. All Site	34 (.14)	147 (.61)	44 (.18)	17 (.07)	242

A similar proportion of classes in each site were of moderate size (10-20 or 21-30 children). However, compared to Detroit and Seattle, Atlanta had a slightly higher proportion of classrooms with fewer than 10 children enrolled and a lower proportion of classrooms with more than 30 children enrolled.

The distribution of classes by enrollment was similar in October and April, except that in Detroit the number of small classrooms decreased by more than 50 percent during the year.

Full-Time and Part-Time Children

Children were enrolled in a center on both a part- or full-time basis. Full-time was defined as an enrollment of 30 or more hours a week on a 4 day or more per week basis; part-time was defined as an enrollment of less than 30 hours per week. Across all three sites, full-time children comprised approximately 70 percent of the total enrollment. The proportions of full-time children in each site were similar (Table 17).

Table 17

FULL-TIME CHILDREN BY SITE, APRIL TARGET CLASSES

	Atlanta	Detroit	Seattle	All Sites
Number of Full-Time Children	1,679	643	850	3,172
Percentage of Full-Time Children	69%	71%	67%	69%

Across all sites, about half of the classrooms had both part-time and full-time children enrolled (Table 18). In Detroit and Seattle, the majority of classrooms had mixed enrollment; however, twice as many Seattle classrooms as Detroit classrooms had exclusively full-time enrollment. The majority of Atlanta classrooms, on the other hand, had only full-time children enrolled. The proportion of part-time children in a classroom was related to the total enrollment: the more children enrolled, the higher the proportion of part-time children was likely to be. In general, target classrooms tended to have a higher proportion of full-time children than classrooms for children under 3 years or over 5 years.

Table 18

FULL-TIME/PART-TIME MIX BY SITE, APRIL TARGET CLASSROOMS

	All Full-Time	Full-Time & Part-Time	All Part-Time
No. of Atlanta Classes (%)	59 (82%)	11 (.15%)	2 (03%)
No. of Detroit Classes (%)	4 (13%)	27 (.87%)	0
No. of Seattle Classes(%)	12 (35%)	21 (.62%)	1 (02%)
All Sites	75 (55%)	59 (.48%)	3 (02%)

The proportion of full-time children enrolled was related to center auspice and funding. Classrooms in public centers and clasrooms in centers that accepted federally funded children tended to have more full-time children than did classrooms in private centers and centers without federally-funded children.

Race and Sex of Children

In the 57 NDCS centers, 62 percent of the children enrolled were black. Of the remaining 38 percent, most were white. Hispanic and Asian children did not make up a significant proportion of the child population.

In Atlanta classrooms, the mean proportion of black children was 78 percent. With the exception of one center, all Atlanta classrooms had a predominantly black enrollment and 67 of the 72 classrooms had only black children. In Detroit, the mean proportion of black children in the classroom was 42 percent. About two-thirds of the classrooms had some black children enrolled, but only four of the 32 classrooms were exclusively black. In Seattle, 74 percent of the classroom population was non-black (including Asian). Black children were enrolled in two-thirds of the 31 classes, but most of these classes had a very small percentage of black children. There were approximately equal numbers of male and female children in the NDCS centers.

Staff and Children

There was a strong correlation between the number of staff assigned to a classroom and the number of children enrolled. However, increases in number of staff were not matched, in general, by increases in enrollment--classrooms with more staff tended to have higher staff/child ratios (Table 19). For instance, across all sites, the single-caregiver classroom had an average enrollment of 15 children. As the number of caregivers increased to two and then to three, the mean class enrollment increased each time by less than two children; thus, the three-caregiver classroom had fewer children/caregiver, on the average, than the two-caregiver classroom.

Table 19

MEAN NUMBER OF CHILDREN ENROLLED AS A FUNCTION OF STAFF SIZE, APRIL TARGET CLASSES

	No. Assigned Staff								
	1	2	3	4	5	6	7	9	11
Atlanta	17.4	12.4	15.4	19.4	26	21.0	23.7	-	-
Detroit	14.9	16.9	15.0	22.0	26	33.3	-	-	42.5
Seattle	18.7	16.9	15.4	21.6	28	-	40.0	36	42.5
All Sites	16.7	14.4	15.4	20.3	26.7	30.2	27.8	36	42.5

For the smaller staff sizes, there was a wide range in the size of the child enrollment (Table 20). For the one-caregiver classroom, the minimum class enrollment was under 10 and the maximum was more than 40. For the two-caregiver classroom, the minimum was less than 10 and the maximum between 30 and 40. In the classrooms with more staff, the range in number of children was smaller.

Table 20

CLASS SIZE BY NUMBER OF STAFF BY SITE, APRIL TARGET CLASSROOMS

ATLANTA

No. of Staff

No. of Children	1	2	3	4	5	6	7
1-10	5	9	3	0	0	0	0
11-15	2	11	8	0	0	0	0
16-20	1	1	6	7	1	0	1
21-30	0	3	4	3	2	1	2
31-40	1	0	0	0	0	0	0
40 or more	1	0	0	0	0	0	0
Mean No. of Children	17.4	12.4	15.4	19.4	26.0	21.0	23.7
No. Classes	10	24	21	10	3	1	3
Total Percent	13.9	33.3	29.2	13.9	4.2	1.4	4.2

DETROIT

No. of Staff

No. of Children	1	2	3	4	5	6	11
1-10	1	0	1	0	0	0	0
11-15	4	4	2	0	0	0	0
16-20	0	7	0	1	0	0	0
21-30	2	1	1	1	1	1	0
31-40	0	0	0	0	0	1	1
40 or more	0	0	0	0	0	1	1
Mean No. of Children	14.9	16.9	15.0	22.0	26.0	33.3	42.5
No. Classes	7	12	4	2	1	3	2
Total Percent	22.6	38.7	12.9	6.5	3.2	9.7	6.5

SEATTLE

No. of Staff

No. of Children	1	2	3	4	5	7	9
1-10	0	3	1	0	0	0	0
11-15	2	6	3	2	0	0	0
16-20	0	2	3	2	0	0	0
21-30	1	3	0	0	1	0	0
31-40	0	1	0	0	1	1	1
40 or more	0	0	0	1	0	0	0
Mean No. of Children	18.7	16.9	15.4	21.6	28.0	40.0	36.0
No. Classes	3	15	7	5	2	1	1
Total Percent	8.8	44.1	20.6	14.7	5.9	2.9	2.9

Although single-caregiver classrooms in Atlanta had the highest mean enrollment among all the sites, the high mean was the result of two atypically large classrooms (see Table 20). The typical single-caregiver classroom in Atlanta had an enrollment of less than 10 children. In Detroit and Seattle, the modal number of children in single-caregiver classrooms was between 10 and 15. For the two-caregiver classes, Atlanta had a lower mean enrollment than Seattle and Detroit. For the classrooms with three or more caregivers, the sites look quite similar in mean enrollment.

In Atlanta and Seattle, the most typical target class size was a classroom with an enrollment of 10-15 children with 2 staff members. In Detroit, the most frequently identified class was one with an enrollment of 16-20 children with 2 staff members. In general, between 60 percent and 75 percent of the classrooms maintained 1-3 staff members with an enrollment of 1-20 children. Classes with 30 or more children were generally staffed by 3 or more people and occurred in less than 25 percent of the total class.

CHAPTER FOUR: CHARACTERISTICS OF DAY CARE CENTER DIRECTORS

Director data from the maximum possible number of centers were included in this analysis, but the selection of a director was not always straightforward. Data collected for five of the 57 centers did not indicate the presence of staff designated as directors, so no one from these centers was included in the present analysis. Multiple directors were identified for five of the remaining 52 centers: one in Atlanta, two in Detroit, and two in Seattle. In the Atlanta center, one individual worked as a full-time director, another worked as a director part-time while working, on a secondary basis, as a teacher. The former individual was specified the director for these analyses. In the first of the two Detroit centers, one director worked 40 hours a week while the other worked only eight, and the full-time director was chosen to represent the center. In the second center, both directors worked only as directors and equally divided the hours of full-time responsibility. Consequently, they were classified as co-directors, and their data were averaged for the center-level analysis. Three individuals were identified as directors in one of the Seattle centers. However, two of these individuals appear to have worked in the center only briefly, so the long-term director alone was included in the analysis. In the other Seattle center, one individual worked as a full-time director while the other part-time director also worked, on a secondary basis, in a supporting staff roll. Again, the individual functioning as a full-time director was chosen to represent the center in analysis.

Director characteristics identified for analysis were previous day care classroom experience, previous day care total experience, previous preschool classroom experience, previous preschool total experience, highest academic degree attained, years of education, hours of total training within the past two years, hours of in-service training within the past two years, current center classroom experience, current center total experience, and presence or absence of educational specialization in a child-related area (see Table 21). A breakdown of director background characteristics by site (APS and non-APS centers in Atlanta; Detroit; and Seattle), by major source of funding, and by auspices (profit, non-profit) revealed only a few significant differences in the distribution of director characteristics.

There were no significant differences across sites for any of the previous day care or preschool experience variables, the training or current center experience measures. However, both of the directors' education variables and the measure of specialization did significantly differ across sites. On the average, APS directors had higher degrees and more years of education, whereas the non-APS directors ranked lowest on both variables. All directors in the APS centers had specialized in a child-related area whereas only about half the directors in each of the other groups had done so.

The qualifications of directors in federally funded centers were not significantly different from those of directors in centers with other sources of funding. However, directors in federally funded centers averaged somewhat more day care experience, education, training, and current center classroom experience. Only preschool experience and total current center experience were recorded as higher among directors in centers without government support. About half the directors in both types of centers had specialized in a child-related area.

There were also no significant differences in the qualifications of profit and non-profit center directors. Directors in for-profit centers had somewhat more day care classroom experience, in-service training, and current center total experience. More of the non-profit center directors had specialized in child-related areas.

Table 21

DIRECTOR QUALIFICATIONS

	APS(N)	Non-APS(N)	Detroit(N)	Seattle(N)	Total(N)
Previous Day Care Classroom Experience	.44(8)	1.22(19)	.24(12)	.98(13)	.81(52)
Previous Day Care Total Experience	.44(8)	4.03(18)	.92(11)	1.45(11)	2.13(48)
Previous Preschool Classroom Experience	4.26(8)	1.62(19)	1.17(12)	1.71(12)	1.95(51)
Previous Preschool Total Experience	4.25(8)	1.67(18)	1.17(12)	3.88(12)	1.98(50)
Highest Degree Attained	3.75(8)	2.17(18)	2.96(12)	2.54(13)	2.70(51)
Years of Education	17.63(8)	14.95(19)	17.00(12)	15.77(13)	16.04(52)
Hours of Training-Last Two Years	67.50(4)	57.92(12)	28.06(9)	53.11(9)	49.87(34)
In-service Training Hours-Last Two Years	134.75(4)	70.58(12)	35.78(9)	115.33(9)	80.76(34)
Current Center Classroom Experience	2.16(8)	.61(19)	.63(12)	.59(13)	.85(52)
Current Center Total Experience	4.25(8)	6.06(19)	5.03(12)	7.30(13)	5.85(52)
Specialization in Child-Related Area	1.00(8)	.50(18)	.55(11)	.42(12)	.57(49)

CHAPTER FIVE: CENTER PHILOSOPHY

Director's Goals

To assess the philosophy of day care held by center directors, seven variables were coded as present or absent from a questionnaire completed by the director: orientation toward individual children; group orientation; cognitive emphasis; ethnic/religious emphasis; guidance emphasis; protective emphasis; and emphasis on behavioral supervision. Each variable is described below. These categories are not mutually exclusive; rather, a center might be categorized as having one or more of these orientations or emphases.

Orientation toward Individual Children

Centers with this orientation stressed the personal development of the individual child. Fostering of independence, self-reliance, learning at one's own pace and self-esteem were key elements in the staff's description of their role with children.

Group Orientation

In these centers stress was placed on the children as members of a group. The focus was on sharing, cooperating, and getting along with peers and adults.

Cognitive Emphasis

This term applied to centers that stressed cognitive development skills such as language learning, number concepts and mastering the alphabet. There was a focus on academic preparation (school readiness) as a program goal.

Ethnic/Religious Emphasis

Centers with this emphasis encouraged the development of a religous or ethnic identity in their children as well as an appreciation of other cultures or religions.

Guidance Emphasis

Staff in centers with This emphasis talked at length about maintaining an atmosphere of love, warmth, security and understanding. They spoke of creating a homelike feeling in the center.

Protective Emphasis

Staff in these centers saw their role as assisting parents by caring for children's basic needs while the parents were working or engaged in other adult activities.

Emphasis on Behavioral Supervision

Stress was placed on obedience, discipline, manners and adult control in centers with this emphasis.

For the 53 centers in the study for which the Program Questionnaires were available, the frequencies of these emphases are shown in Table 22. Most centers emphasized cognitive development and tried to maintain a warm, homelike environment. All other emphases were held in about one quarter to one third of the centers.

The patterns of the contingency coefficents shown in Table 23 indicate that there are two basic types of centers. The first has a group orientation which is accompanied by emphases on child obedience and the protection of children while their parents are away. This kind of center

Table 22

FREQUENCIES OF OCCURRENCE OF THE PHILOSOPHY VARIABLES

Emphasis	No. of centers	% of centers
Group	21	39.6
Individual	13	24.5
Cognitive	42	79.2
Ethnic/Religious	12	22.6
Guidance	36	67.9
Protection	13	24.5
Supervision	14	26.4
Total	53	

Table 23

INTERCORRELATIONS OF PHILOSOPHY VARIABLES[1]

	Indi.	Cog.	Eth./Rel.	Guid.	Pro.	Super.
Group	-.419**	-.404**	--	-.332*	.399**	.363*
Individual	--	--	--	--	--	--
Cognitive		--	--	.327*	-.422**	-.310*
Ethnic/ Religious			--	--	--	--
Guidance				--	-.638***	-.567***
Protection					--	.484***

---

[1] All correlations are contingency coefficients.

*   p<.05

** p<.01

*** p<.001

does not emphasize the individual, nor does it stress cognitive development or the maintenance of a warm, homelike atmosphere. The second kind of center does emphasize the learning of cognitive skills in a loving atmosphere and does not see its role in a guidance or protection framework. The first kind of center seems therefore to be maintaining a safe environment in which children can play while the second kind is providing a comfortable learning center for children.

Classroom Goals

Lead teachers were also asked to indicate the orientation and goals of their classrooms. First, they chose between a learning atmosphere and a homey atmosphere; second, they chose one of four goals for their class: academic preparation, getting along with other children, getting along with teachers, taking care of oneself. Across the three sites, more classrooms were designated as striving for a homey atmosphere than a learning atmosphere (46% vs. 35%; 18% selected neither). In Atlanta and Seattle, this same pattern held. In Detroit, however, 48 percent of the lead teachers selected "learning atmosphere," compared with 29 percent who selected "homey atmosphere". Whether a classroom was reported to have a learning atmosphere did not predict the amount of time spent daily in group learning activities. Classrooms in both categories spent an equal amount of time--a little more than one hour daily--in group learning activities.

The most frequently selected classroom goal in all sites (coded as "1") was "getting along with other children" (Table 24). The least selected goal was "getting along with

Table 24

ORDER OF IMPORTANCE OF CLASSROOM GOALS, AS REPORTED BY LEAD TEACHERS IN APRIL TARGET CLASSES

	Academic Preparation	Get Along W/Children	Get Along W/Teachers	Take Care of Oneself
Atlanta	3	1	4	2
Detroit	2	1	4	3
Seattle	3	1	4	2

teachers." In Detroit, "academic preparation" was the second most frequent goal, while "taking care of oneself" was second in Atlanta and Seattle.

Classes which indicated an orientation toward academic preparation spent an average of one hour and twelve minutes a day in group learning activities; centers which were not academically oriented spent significantly less time--an average of only 54 minutes a day. Classrooms whose primary goal was academic preparation more often reported having cognitive learning activities (i.e., school readiness) than classrooms oriented toward social or personal skills.

CHAPTER SIX: MEASURES OF SPACE AND THE PHYSICAL ENVIRONMENT

Measures taken from center space plans, the Physical Environment Inventory (PEI) section of the Adult-Focus Instrument, and the Child Development Associates (CDA) Checklist were included as indices of the physical environment, and are discussed in this section.

Center Space Plans

Several variables were derived from the information on center space plans. Two were class-level variables: number of square feet in target class's HOMEROOM; and amount of center space in daily use by the target class (CLASS SPACE). Three could be considered center-level or class-level variables, but because their values are identical for all classes in a center, are best thought of as center-level indices: amount of space in the center used daily by any or all of the target classes (CHILD SPACE); all center space dedicated to day care use (INDOOR SPACE), equal to the sum of CHILD SPACE and space used for storage and adult activities; and amount of outdoor play space designated for day care children (OUTDOOR SPACE).

Table 25 gives the mean square footage by site for each of these variables. The size of HOMEROOMs in Detroit tends to be larger than that in Atlanta or Seattle, and this difference holds for CLASS SPACE and for CHILD SPACE. The INDOOR SPACE set aside for day care activities in Seattle is almost as large as that in Detroit due to the more spacious facilities for adults and for storage in Seattle. With OUTDOOR SPACE, Atlanta centers appear to be more generous. Taken in concert with the larger size of CHILD SPACE in Detroit, it would seem as though these findings reflect the relative use of indoor and outdoor space in the two climates. In Atlanta where there are many more days in which outdoor play is possible, centers have more space for such play. In

Detroit where the weather is inclement for many months of the year, resources have been directed more at indoor facilities. Seattle falls in between the two extremes.

Table 25

MEASURES OF SPACE (SQUARE FOOTAGE) BY SITE

	Atlanta	Detroit	Seattle
Class Level	N=56	N=28	N=36
Homeroom	753	1,049	736
Class Space	2,264	4,116	3,162
Center Level	N=23	N=12	N=16
Child Space	3,543	4,361	3,695
Indoor Space	4,187*	4,975	4,870
Outdoor Space	7,709	5,295*	5,912

---

*One center has been removed from the calculation of each of these means because of its extreme values.

The Physical Environment Inventory

The Physical Environment Inventory (PEI) provides a record of the equipment available in the classroom and requires that an observer note any space and equipment problems and the degree to which children can choose the space in which they will play. The first 11 categories on the PEI describe the soft materials in the classroom environment (cozy furniture, carpet or rug, grass to be on, sand to play in, dirt to dig in, animals to hold, sling swings, dough, messy materials such as fingerpaints, water for play, and stuffed toys). A classroom with a high total score on soft materials has many flexible, malleable materials which children can touch. The second section of the PEI records the pressence of 15 kinds of play equipment (dramatic play props, games and puzzles, books, water play, vehicles to ride, dolls, outdoor physical equipment, construction table toys, work tools, floor blocks, small wheel toys, skill equipment for carpentry, cooking, etc., art materials, swimming pool, child-sized furniture). A classroom with a high score on play materials offers many different materials for children's use. The last sections ask that observers judge the existence of space and equipment problems and the degree of choice children have in selecting activities and space in the classroom.

Multiple Inventories were completed on a single classroom over two days of observation. Typically, two inventories were recorded on a class per day. Some classes, however, had as few as one Inventory, and others had as many as 10, depending on the number of days of observation and the number of caregivers observed.

In the spring data collection, several soft materials were present in most classrooms. Sand was available to 90 percent of the 134 classrooms, messy materials to 87 percent, carpeting or rugs to 84 percent, grass and dirt to 78 percent, stuffed toys to 76 percent, sling swings to 72 percent, and dough to 66 percent. Three of the categories of soft materials were relatively uncommon. Cozy furniture was only present in 31 percent of the classrooms, water play in 30 percent, and animals to hold in 22 percent.

By summing across all 11 categories of soft materials, it was possible to create a softness index for each classroom. The mean scores of this index for the three sites are displayed in Table 26. The scores are all about seven, indicating that most classrooms have a good number of soft materials with which children can play. Seattle classrooms tended to have soft materials more often. Site differences were significant for cozy furniture, carpet, dough and water, though they were not significant on the overall Softness Index.

Table 26

MEASURES FROM THE PHYSICAL ENVIRONMENT INVENTORY BY SITE

	Atlanta	Detroit	Seattle
	(N=70)	(N=29)	(N=35)
Softness Index	6.9	7.1	7.7
Kinds of Play Equipment	11.6	12.2	12.1
No. of Space Problems	1.2	2.3	1.2
Space Selection	1.9	2.2	2.2

Of the 15 kinds of play equipment, many were observed in virtually all of the classrooms: games and puzzles (99%); art materials (99%); child-sized furniture (99%); books (98%); outdoor physical equipment (98%); construction table toys (98%); floor blocks (96%); small wheel toys (96%); dramatic play props (94%); and dolls (92%). Vehicles to ride were usually present (78% of the time). Other materials were present in less than half of the classrooms: skill equipment (46%); water play (38%); work tools (37%); and a swimming pool (15%). Most of the classrooms were thus well provided with equipment appropriate to the age of the children.

A total equipment score was calculated for each classroom by summing across all 15 categories. As shown in Table 26, classrooms in all sites averaged about 12 kinds of play equipment. No classroom had an equipment score less than 5; about 10 percent of the classrooms had equipment scores of 13 or more, suggesting a wide range of types of equipment. A few significant site differences existed on individual kinds of equipment, but none of the sites was consistently better equipped.

Six space or play equipment problems could be noted on the PEI, and a total score from 0 to 6 was recorded from the noted problems. These problems included lack of shade, broken or shabby equipment, space used as a pathway for other people, two groups in one space interfering with each other, little privacy for children, all asphalt or rough ground. None of these problems as observed in about one-quarter of the classrooms; one-third of the classrooms had one problem only; only 10 percent of the classrooms had three or more problems. The means shown on Table 26 demonstrate that classrooms in Detroit had more problems than those in Atlanta or Seattle, but on an absolute scale, still had relatively few problems.

Observers coded the degree of choice children had with space on a scale from 1 to 3 where 1 meant no choice, 2 some choice, and 3 high choice. Seventy-three percent of the classrooms were coded as providing some choice. The means for this variable in all sites are about 2, though Atlanta classrooms offered children significantly less choice in their activities than classrooms in the other sites.

Factors from the CDA Checklist

The CDA Checklist was administered in conjunction with Adult-Focus Observations to provide additional information about caregiver skills. The checklist was developed to reflect 11 functional areas that constitute a basis for awarding the CDA credential.* To simplify analyses of the checklist data, items were factor-analyzed within functional areas with 12 resulting factors. Four of the factors--safety, health, physical competence materials and creative materials--serve as indicators of the classroom physical environment. Site-specific means for these four factors are displayed in Table 27.

---

*A complete description of this instrument is presented in N.N. Goodrich, "An Analysis of the CDA Checklist Data." In National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, MA: Abt Associates Inc., 1980.

27

Table 27

ENVIRONMENTAL MEASURES FROM THE CDA CHECKLIST BY SITE

	Atlanta (N=70)	Detroit (N=30)	Seattle (N=34)
Safety	.94	.91	.95
Health	.94	.89	.94
Physical Competence Materials	.49	.52	.61
Creative Materials	.57	.81	.82

Each of the numbers in Table 27 indicates the mean proportion of items within the factor that were observed in the classroom. That is, in Atlanta, an average across centers of 94 percent of the items relating to safety were observed; a mean of 94 percent of the items pertaining to sanitation were noted; and an average of 49 percent of relevant physical competence materials and 57 percent of the relevant creative materials were present in centers.

It is clear from the table that most of the centers were safe and sanitary, and that the sites do not differ on these variables. However, the sites do differ somewhat on the presence of physical competence and creative materials, with Atlanta relatively less well equipped and Seattle somewhat better-equipped. These trends are in the same direction as those on the PEI.

Creating an Index of the Physical Environment

In searching for clusters of variables which might provide valuable indices of differing physical environments, factor analyses were run on the center space variables in conjunction with four PEI variables (softness index, number of kinds of play equipment, number of space problems, and amount of freedom allowed in selecting play area) and four variables from the CDA Checklist (safety, health (sanitation), number of kinds of physical competence equipment present, and number of kinds of available creative materials). The factor loadings for one typical analysis are presented in Table 28. Two of the factors (1 and 4) have high loadings for variables derived from the space plans; Factor 2 represents the safety and health measures in the PEI and CDA Checklist; and Factor 3 brings together measures of kinds of equipment in the centers. Thus, there is no single clustering of the physical environment measures which would be useful in classifying space as "good" or "bad." Rather, the measures of amount of space are quite separate from measures of the quality of the use of the space.

Table 28

FACTOR LOADINGS FOR ENVIRONMENTAL VARIABLES*

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Softness Index (PEI)	.19146	.12558	.45303	.05175
Kinds of Equip. (PEI)	.13568	.08359	.46074	.06756
Space Problems (PEI)	.10171	-.89632	-.06684	-.25213
Space Selection (PEI)	-.11704	.03855	.51019	.11298
Size of Homeroom	.16140	-.20672	.16394	.81650
Class Space	.90249	.02015	.28661	-.12350
Child Space	.97102	.08022	.11266	.07999
Indoor Space	.91071	.15847	-.04325	.30727
Outdoor Space	.09227	.06413	.12561	.21566
Safety Factor (CDA)	.19818	.37510	.03789	-.03047
Health (Sanitation) (CDA)	-.02400	.41454	.11079	-.02446
Physical Competence Materials (CDA)	-.04346	.11130	.01228	.24194
Creative Materials (CDA)	.05748	.04252	.60733	.05425

---

*A Rao Factor Analysis was performed with a varimax rotation. Similar results were reached with a principal components analysis.

Summary

The purpose of this paper was to provide descriptive information on the classrooms and centers in the NDCS for those readers who would like to know the context in which this study of child care was conducted. For further information on the relationships of these variables to outcomes of interest, the reader is referred to other papers in this volume, especially J.D. Singer, "Classroom Process--Child Outcome Analysis."

The Econometric Model

Daniel Calore

TABLE OF CONTENTS

Page

CHAPTER ONE: INTRODUCTION 263

CHAPTER TWO: ECONOMETRIC MODEL OVERVIEW 265

- Variables in the Model 267
- Major Behavioral Hypotheses 268

CHAPTER THREE: OVERVIEW OF ECONOMETRIC RESULTS 275

CHAPTER FOUR: MODEL SIMULATION 284

- Methodology 284
- Baseline Evaluation 287
- Simulation Results 289

CHAPTER FIVE: CONCLUSION 299

APPENDIX A 303

APPENDIX B 309

APPENDIX C 315

CHAPTER ONE: INTRODUCTION

The expanded influence of government at all levels in the market for child care services brings with it an increased governmental responsibility to understand and control the effects of that involvement. The likelihood that the full impact of policy changes will be unanticipated or go unrecognized rises with the extent of this involvement and can introduce additional complexities to the market environment and social policy. A need for improved analytic devices which will predice the full implications of policy changes has been recognized by policymakers and planners in the fields of transportation, energy, agriculture and health to name a few. The simulation model presented here is an analytic tool which will help satisfy this need in the field of child care.

The model is designed as a planning tool which has the ability to accept a very broad range of policy configurations. Among the controllable policy variables contained are federal child/staff ratio requirements, federal group size restrictions, state reimbursement rates, and frequency of monitoring visits. With some minor adjustments to the data set, the child/staff ratio requirements of individual states or groups of states can be adjusted and local impacts assessed.

All of this can be accomplished with the mathematical technique of simulation. The approach is conceptually simple, and the results are straightforward and easily interpreted by anyone familiar with the day care industry. Both as a demonstration of the model's power and for the practical purposes of evaluating current federal policy alternatives, the following sections explain and then apply the simulation model to ten different policy experiments.

Chapter Two gives a brief economic background to the day care industry and presents some of the hypotheses which the model will test. Estimation of the model--the derivation of its structural parameters--is described in Chapter Three and the most important of these are interpreted to provide some insights into the market environment of day care. Chapter Four presents the results of ten simulated policies and compares them to current market conditions as measured by forty-six variables.

CHAPTER TWO: ECONOMETRIC MODEL OVERVIEW

This paper evaluates the impact of regulatory policy upon the day care industry by means of a multi-equation econometric model. In this model, demand is assumed to originate from two sources: a private sector which includes parental funding, private charities and scholarships; and the welfare and community social service programs of the public sector.

Day care services are supplied by an assortment of firms varying in size, revenue source, and organizational structure which fall into four categories:

- Private profitmaking centers,
- Private non-profit centers,
- FFP* profitmaking centers, and
- FFP non-profit centers.

The model explains the behavior of these types of day care centers as they respond to their market environment represented by demand relations for care, variables describing local characterisitcs, and a series of parameters, variables and equations which characterize the set of state and federal regulations.

The sample is cross-sectional; that is, many centers are observed during a single common period.** The

---

*The term FFP is an abbreviation for "federal financial participation." Strictly defined, the term "FFP center" denotes a center which has at least one child enrolled whose care is paid for under Titles IV and XX of the Social Security Act and miscellaneous other federal programs; see Coelen, Glantz, and Calore, Day Care Centers in the U.S.A.--A National Profile, 1976-1977, Abt Associates Inc., National Day Care Study, Vol. II, 1978.

**This is not strictly so, because the surveys from which the data are taken were conducted in four seasonal rounds; see Coelen, Glantz, and Calore (1978).

28-

benefits of such a sample stem from its automatic controls upon time-trend factors such as changes in educational technology, national policy, and social forces such as the expanding opportunities for women. The cross-sectional sample will not hold interregional and local differences constant, however. For example, labor market conditions, local prices, demographic characteristics, and state and local day care policy are not controlled by the sample. Instead, a series of local and regional dummy variables have been constructed for use as adjustments to the regression equations where needed.

The model describes a time period of intermediate duration: a span long enough to allow all costs including so-called fixed costs, such as rent, to vary, but yet not so long that major operating characteristics of the center, like legal status and licensed capacity, will change.

The system is constructed to replicate the multidimensional response to policy change through the simultaneous influences of variables across the set of equations. The actual outcomes of the simulations will be comparative static solutions for the endogenous* variables of the model.

---

*The terms "endogenous," "exogenous," and "predetermined variables" will be used throughout this paper. The three types of variables are distinguished primarily by the point at which their values are determined. Values of endogenous variables depend upon the forces acting upon the model both internally and from outside. Thus they change subject to changes in other variables in the model. Exogenous variables are unaffected by the workings of the model. Their values usually reflect the conditions of the economy at large, demographic characteristics or technology. Changes in exogenous variables used in the model have an effect on some or all endogenous variables.

A hybrid of the two types--"predetermined variables"--includes all exogenous variables and all endogenous variables whose values are determined in some period prior to that for which analysis is performed. Examples of the latter group are lagged values of center enrollment and a variable describing the center's legal status.

This differs from the typical time series simulation which permits measures of speeds of adjustment to external shocks. This difference will not affect the model's ability to describe the new equilibrium following a policy change, however, since the order of impact and response is preserved.

Variables in the Model

Appendix A gives a list of the variable names and their definitions as used throughout this paper. The number of predetermined endogenous variables in the model is small; they are primarily used to specify factors held constant throughout the period of analysis (e.g., legal status). Others are dummy variables for donated space and group or center size. Exogenous variables specify local or regional characteristics. Among these are state median family income, state mean parent fees, and a dummy variable for urban location. To control for regional variability in federal day care policy, nine dummy variables for Department of Health, Education and Welfare regions 1 to 9 have been added.*

Most of the endogenous variables are linked to one another in complex interdependencies. All staffing, service, cost, and enrollment variables (other than number of government-funded children) are treated as if they were determined simultaneously. Regulatable characteristics such as group size, child/staff ratios, caregiver characteristics, and the impact variables are also handled with this group. Since government enrollment is determined before the values of these variables are computed, the size of government enrollment influences, but is not influenced by, the values of all other endogenous variables in the system.

---

*HEW regions 4, 7, and 8 were not significant in the model estimation.

A schematic description of the general relationship present in the model is given in Figure 1. Total enrollment, consisting of privately and publicly funded children and governmental standards of care prescribed for the size and age composition of that enrollment combine to determine the size and occupational characteristics of the center's staff. Enrollment and regulations will also influence the nature and extent of supplemental services provided by the center. Staffing, services, and enrollment affect cost directly though to very different degrees. Fees paid by private sources are determined largely by the level of average cost and, in turn, have a strong impact on demand for care by the private sector.

All endogenous variables fall into one of six broad categories. The most important variables are defined in Table 1. All are listed and defined in Appendix A.

Major Behavioral Hypotheses

An understanding of the properties of several market forms may be necessary to explain the diversity of behavior shown in the sample. For example, non-profit centers, both private and FFP, clearly use different decision rules when adjusting to changing market conditions than do the profitmaking centers. The simple response to a general rise in the demand for care will likely be an increase in fees by profitmakers. Non-profit centers, if motivated by a desire to maximize the number of children served, would probably increase capacity and not necessarily raise fees.

Centers differ somewhat from one another in the number and quality of services they offer, their size, fee structures, and staffing characteristics with the greatest differences occurring between FFP and non-FFP centers. The factors are sufficient, however, to prevent parents from seeking care elsewhere if price rises too far. In other

Figure 1

GENERAL RELATIONSHIPS FOR ECONOMETRIC MODEL

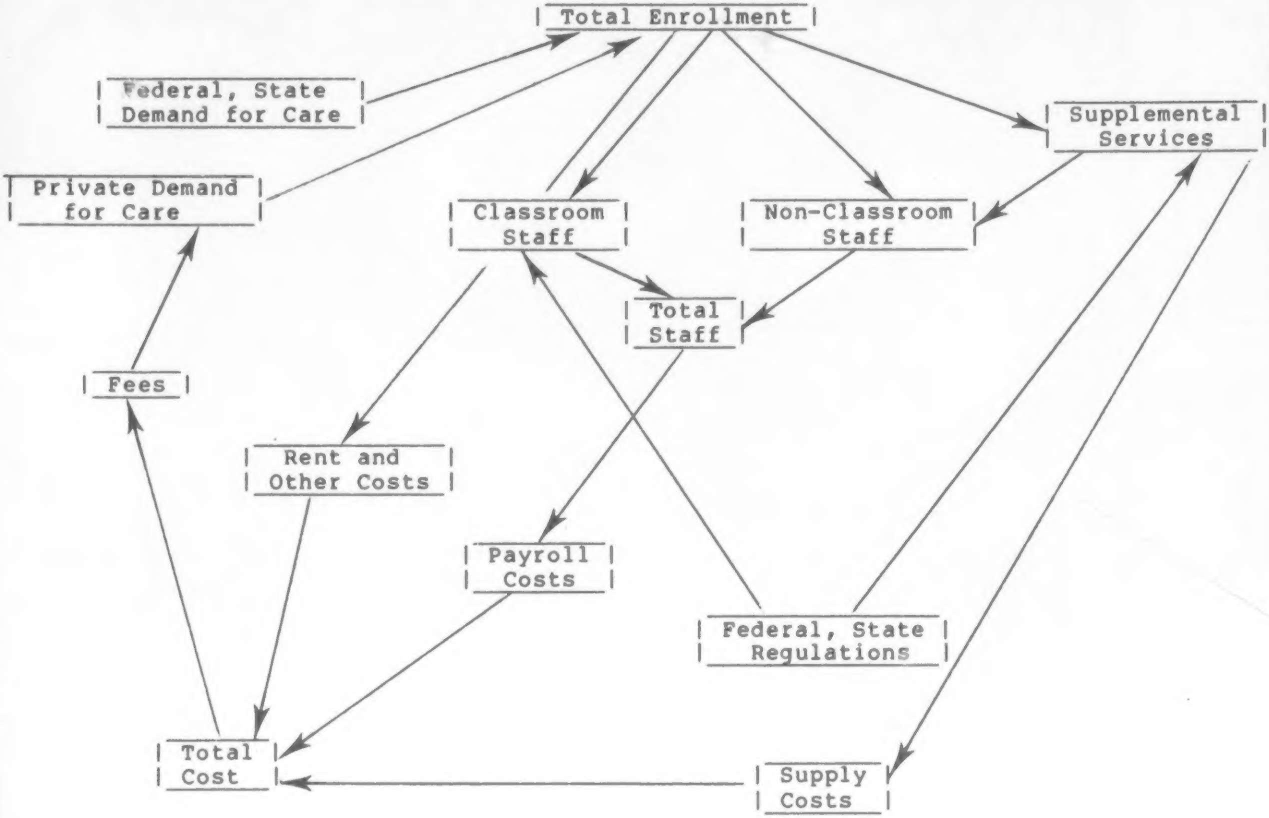
Total Enrollment
Federal, State Demand for Care
Private Demand for Care
Supplemental Services
Classroom Staff
Non-Classroom Staff
Total Staff
Fees
Rent and Other Costs
Payroll Costs
Federal, State Regulations
Total Cost
Supply Costs

Table 1

1. General Center Characteristics

- Enrollment

FRNO = headcount enrollment of privately funded children

GOVKIDS = headcount enrollment of government-funded children

ENR = total headcount enrollment

ENR = FRNO + GOVKIDS

KFTE = full-time equivalent enrollment

- Socioeconomic Mix

RICHB = percent of families served with household income over $15,000 per year

POORB = percent of families served with household income of $6,000 per year or less.

2. Program Characteristics

- Services Offered

NMEALS = number of meals served to full-time children per week

TRANSP = number of children for whom transportation is provided

PEX = a dummy variable equal to 1 if the center offers physical exams, 0 otherwise

DEX = a dummy variable equal to 1 if the center offers dental exams, 0 otherwise

PARSERV3 = a dummy variable equal to 1 if the center offers assistance to parents in obtaining general financial aid, 0 otherwise

- Participation of Parents in Center Activities

PART = a dummy variable equal to 1 of parents participate in staff selection, 0 otherwise

REVIEW = a dummy variable equal to 1 if parents review center programs and budgets, 0 otherwise

- Classroom Characteristics

ACSR = actual child to staff ratio

AVSIZE = average group size of groups containing three- or four-year old children.

CFTE = number of full-time equivalent caregivers.

Table 1 (cont.)

3. Staff Characteristics

EDMEAN = average number of years of education of caregivers

EXMEAN = average number of years of day care experience of caregivers

NCARE = headcount number of non-caregiving staff

FTEPDSTF = full-time equivalent paid staff

VOLSTAF = headcount number of volunteers and staff paid by outside agencies

SW = a dummy variable equal to 1 if center has a social worker on staff, 0 otherwise

NURSE = a dummy variable equal to 1 if center had a nurse on staff, 0 otherwise

4. Financial Characteristics

AVESAL1 = average monthly staff salary

MONRENT = monthly center expenditure for rent

MONSUP = monthly center expenditure for supplies

SALEXP = monthly center expenditure for payroll

OTHCOST = monthly center expenditure for miscellaneous items

MONCOST1 = total monthly center expenditures

MONCOST1 = SALEXP + MONRENT + MONSUP + OTHCOST

COSTPK = average cost per FTE child

COSTPK = MONCOST1/KFTE

5. Regulatory Variables

REQGRP = number of FIDCR-required groups

FRSHRS = FIDCR-required staff hours

SRSHRS = state-required staff hours

FCSR = FIDCR-required child/staff ratio

FCSR = 40*KFTE/FRSHRS

SCSR = state-required staff hours

SCSR = 40*KFTE/SRSHRS

6. Compliance Measures

FOK = a dummy variable equal to 1 if the center complies with current FIDCR staff/child ratio requirements, 0 otherwise

SOK = a dummy variable equal to 1 if the center complies with state staff/child ratio requirements, 0 otherwise

AMFR = actual minus FIDCR-required FTE staff members

AMSR = actual minus state-required FTE staff members words, centers must compete on the basis of fees to maintain their private enrollments. The result is a narrow distribution of fees charged to parents for care.*

Although FFP centers engage in competition** over fees, the characteristics of the total demand they face are significantly different from that in private centers. The private component of total demand has the same characteristics as that in private centers. The government component, however, is considerably less responsive to price changes for a number of reasons. For example, government programs usually mandate supplementary services which the private sector purchases less frequently. Thus, price is only one of several factors considered when the government purchases care and is, therefore, a comparatively less important factor to the government in setting the quality demand.

One maintained hypothesis at work in this model is the autonomy of government enrollment in center operations. The variable is specified as a function of predetermined variables only and so is not affected by internal reactions of the model. The implication is that the governmental agency which places children is obliged to find slots regardless of market conditions, although the number of placements sought may vary with price due to some budget constraint. The number of government-funded children then becomes a very important variable in determining the size of privately funded enrollment, the number of FTE caregivers, the number of noncaregivers, and the age distribution of enrollment.

---

* Over one-half of all centers charged maximum fees between $21 and $30 per week to their parent-paid enrollment in 1976-1977 (Coelen, Glantz, Calore, 1978).

**None of the competitive behavior discussed need be very aggressive. It can take hte simple form of center directors resisting fee increases above some widely accepted limit.

The "a priori" notions about the labor supply of caregivers must be based upon observations of the market for teachers to some extent. This is necessary because of the lack of developed data sources for day care labor and because of distinction in skills between primary grade teachers and center caregivers is not sharp enough to prevent spillovers from one labor force to the other. The current excess supply of teachers is nationwide in scope and of apparently long duration. Therefore, caregivers act in a highly competitive market and individually have little effect upon the wage they command. On the other hand, demand among centers for caregiver services is somewhat competitive because centers are numerous and so must contend, to some extent, with caregiver demand by grade schools, kindergartens, family day homes, and alternative occupations. The wage level that prevails, then, should be determined jointly by these characteristics of caregiver demand and supply.

Because the demand for caregivers is strongly dependent upon federal staffing requirements in FFP centers and perhaps even state requirements in all centers, the need for regulatory compliance will mute the effect of wages on caregiver staff size even in spite of a general noncompliance with these regulations.* The size of the caregiving staff varies directly with the presence of volunteers because centers can add non-paid caregivers at little or no extra cost. This can be done in response to regulation or to some change in center operating policy.

Other staffing decisions are affected only indirectly by regulations. The number of center staff other than caregivers depends upon the amount and quality of supplementary services provided by the center. Education and experience levels of the staff reflect the characteristics of the local labor force as much as individual center needs. In particular, educational attainment among caregivers is higher in more affluent urban, better educated populations.

---

*It is possible for a law to influence behavior even in the absence of strict compliance with it. Imposition of a 55 mile-per-hour speed limit alone reduced average highway speed but did not bring it down to 55.

The provision of supplementary services does not depend upon usual supply considerations such as the cost of inputs or market prices. Most are either mandated by regulations or provided routinely by non-profit centers which frequently have services donated or provided at reduced cost. Furthermore, empirical results show that total costs are not affected by supplemental services in a substantial way. The major exceptions are the provisions of meals and transportation, the costs of which vary with enrollment and usually require other expenditures such as those for kitchen facilities and additional staff.

On the other hand, total cost does depend upon the scale of center operation and so will vary with enrollment as well as the age composition of enrollment. Since child/staff ratio requirements vary with age, the age distribution of enrollment will affect staff size and expenditures on special supplies for very young children. All components of cost will naturally be very sensitive to local price levels, though each to a different degree.

CHAPTER THREE: OVERVIEW OF ECONOMETRIC RESULTS

This chapter presents the findings obtained from equation estimations carried out prior to the simulation experiments.*

Results of all equation estimations appear in Appendix C. A selected number, primarily those for enrollment, staffing, and costs, will be discussed here. The final functional forms differ in three important ways from original specifications. First, distinctions made between profit and non-profit and between FFP and private centers were of much less significance in staffing equations than was expected earlier. In many cases, structural differences between center types disappear when enrollment of welfare children is held constant. Other continuous variables possess distributions which already reflect the distinction between profit and non-profit centers: an obvious example is the equation for the number of volunteers, but others, such as those for private enrollment and FTE caregivers, have this property as well, though to a smaller degree. In other cases the distinction among center types simply does not exist and equations to explain these, such as those for staff education and monthly supply expense, do not benefit from the added variables.

A second general finding is that service variables, except for the number of meals, play a very limited and indirect role in total cost determination. Hearing and vision testing and psychological testing were of no significance; physical and dental exams given by centers influenced costs only where a nurse was present on the center's staff. The nurse increased the number of noncaregivers by one-half person, which affected payrolls. Only the presence of a staff social worker was of comparable importance. Financial advice and referral is the only service to parents with significant effects on other variables, in particular on the presence of a social worker. In this case, some question arises about the causal direction. Does a center decide to offer referral services and then hire a social worker to provide the advice, or are the services offered because the social worker is available? The former seems more likely, but without additional information not presently available, the uncertainty cannot be resolved.

---

*A discussion of the data and methodology which might logically precede this section has been placed in Appendix B. This was done to preserve the continuity of the present topic while devoting necessary attention to important econometric issues.

Finally, of those variables whose values are within the control of center directors (called decision variables), the variable most sensitive to changes in cost is the number of parent-fee children enrolled. Parent-fee enrollment, however, has important and widespread effects on staffing, total enrollment, the age distribution, and the socioeconomic mix of the center and thus is a principal feedback route into the system from cost changes.

Results for Equation (1) show residual capacity (CAP - GOVKIDS) as a significant determinant of parent-fee enrollment. A unit increase in residual capacity in FFP centers raises parent-paid enrollment by 0.7 children. A one-tailed t-test shows this coefficient to be significantly less than 1.0, which indicates that in shifting from government-funded to parent-funded children, centers do not do so in a one-for-one substitution. Instead, on the margin, ten government-funded children would be replaced by only seven parent-paid children. This implies that centers tend to make greater space, staffing, and/or service accommodations for privately funded than for government-funded children.

1. $$\text{FRNO} = \underset{(4.72)}{-9.27} + (\underset{(0.0227)}{0.687} + \underset{(0.0269)}{0.08}*\text{LEGSTAT})*(1-\text{FFP})*\text{CAP}$$

$$+ \underset{(0.0274)}{0.7}*\text{FFP}*(\text{CAP} - \text{GOVKIDS}) + \underset{(0.000519)}{0.0033}*\text{MEDY} - \underset{(0.013)}{0.109}*\text{COSTPK}$$

$$- \underset{(1.52)}{3.48}*\text{ZIPDUM4}$$

$$R^2 = 0.699 \quad F(6,732)=284.$$

$$\text{SER/LHS MEAN}=18.5/40.5=0.457$$

Where:

LEGSTAT = A dummy variable equal to 1 if the center is a profit center, equal to 0 if it is not.

CAP = Licensed capacity of center.

FFP = A dummy variable equal to 1 if the center is a FFP center, equal to 0 if it is not.

MEDY = Median annual family income in center catchment area

ZIPDUM4 = A dummy variable equal to 1 if the center is located in a SMSA, equal to 0 otherwise.

The ratio of welfare reimbursements to average parent fee is significant and positive in Equation (2), whereas the absolute reimbursement rate was not significant in earlier specifications. The result is intuitive--as the reimbursement rate rises relative to the fees charged to parents, day care services to welfare children become an increasingly better market. Equation (9) gives full-time-equivalent enrollment as a linear combination of headcount enrollments by funding source. The reciprocals of the coefficients for each of the enrollment variables give the comparative rates of change in the two groups as full-time enrollment changes. For example, a one FTE child increase in enrollment raises the number of parent-paid children in non-FFP centers by 1.21, in FFP centers by 1.28, or the number of welfare children by 1.02.

2. $$\text{GOVKIDS} = \text{FFPA}*\text{CAP}*(\underset{(0.0339)}{0.00119} + \underset{(0.0277)}{0.185}*\text{WELFIX2}/\text{FEEBAR}$$

$$+ \underset{(0.96)}{7.15}/\text{CAP} - \underset{(0.0179)}{0.137}*\text{LEGSTAT} * \underset{(2.45)}{4.25}*\text{FTCOST}/1{,}000{,}000)$$

$$R^2 = 0.522 \qquad F(4,734)=200.0$$

$$\text{SER/LHS MEAN}=8.96/6.25=1.43$$

Where:

WELFIX2 = Weekly welfare fee per child.

FEEBAR = State average parent fee per child.

FTCOST = Total Title XX expenditures for full-time placements in center care.

9. $$\text{KFTE} = \underset{(3.57)}{-51.2} + \underset{(0.337)}{4.91}*\text{HRSOP} + \underset{(0.696)}{2.64}*\text{SUMDUM} + (\underset{(0.0105)}{0.826} - \underset{(0.0146)}{0.042}*\text{FFP})$$

$$*\text{FRNO} + \underset{(0.0267)}{0.977}*\text{GOVKIDS}$$

$$R^2 = 0.918 \qquad F(5,733)=1640$$

$$\text{SER/LHS MEAN}=8.53/41.3=0.207$$

Where:

SUMDUM = A dummy variable equal 1 if center was interviewed for the study during the summer of 1976.

HRSOP = Number of hours per day the center is open.

Among the important results in the staffing equations is that salary is not a statistically important factor in setting caregiver staff size or total center staff size.* Two possible explanations are:

- Salaries are determined only after a sequence of decisions are made including enrollments, staff size, reimbursement rates, etc.

---

*The widely accepted neoclassical microeconomic optimization models of the firm describe an economically efficient production process as one where each input is used up to the point where its contribution to output is just equal to its addition to cost. This condition is not fulfilled according to our analysis.

- The role of salaries in determining staff size is overpowered by the need to satisfy ratio requirements.

Centers employing a nurse or social worker have roughly one-half a noncaregiver more than centers that do not employ these professionals. This is true whether the professional is paid or acts as a volunteer. Each additional ten volunteers adds 3.1 noncaregivers. Of each additional 100 hours of staff time in all centers, 84 hours are paid.

The cost regressions underwent substantial changes from their original specifications. In particular, many of the controlling dummies for center type and region were dropped in the salary and rent equations.

Equation (11) estimates average monthly salary as a function of the caregiver-to-total staff ratio on the assumption that the closer its value is to unity, the greater is the upward pressure on wages. This may reflect a substitution of expenditures on caregivers for expenditures on supplementary services--the latter represented by the number of noncaregivers, who are paid less per person on average as a group. Average caregiver salary rises by $51 per month as the ratio of caregivers-to-total paid staff rises from zero to one. This ratio can be greater than one because FTE caregivers may include volunteers. Also, the higher the ratio of government-funded children to total enrollment is, the easier a center will find it to pass on increased labor costs.

The average educational attainment and years of experience of caregivers each have a positive effect upon salary levels. However, the effect of an additional year of education is substantially greater than that of an additional year of experience. The average monthly salary for caregivers in HEW Region 2, the Mid-Atlantic states, is about $62 higher than the national average when other factors are held constant.

11.

$$\mathrm{AVESAL1} = \underset{(46.2)}{58.7} + \underset{(11.9)}{51.0} * \mathrm{CFTE/FTEPDSTF} - \underset{(12.2)}{28.3} * \mathrm{FFP} * \mathrm{LEGSTAT}$$

$$+ \underset{(0.0031)}{0.0179} * \mathrm{MEDY} + \underset{(3.2)}{13.4} * \mathrm{EDMEAN} + \underset{(19.3)}{61.7} * \mathrm{REG2} + \underset{(1.14)}{2.04} * \mathrm{EXMEAN}$$

$$+ \underset{(0.659)}{1.41} * \mathrm{DGSZ} * \mathrm{AVSIZE} + \underset{(15.2)}{80.4} * \mathrm{GOVKIDS/ENR}$$

$R^2 = 0.186 \quad F(8,730)=20.9$

SER/LHS MEAN=106./496.=0.214

Where:

REG2 = A dummy variable, equal to 1 if the center lies within HEW's Region 2, and equal to 0 otherwise.

DGSZ = A dummy variable, equal to 1 if average center group size is less than 40, and equal to 0 otherwise.

3.

$$\mathrm{CFTE} = \underset{(0.139)}{0.798} + (\underset{(0.000814)}{0.00955} + \underset{(0.000539)}{0.00487} * \mathrm{FFP}) * \mathrm{FRSHRS}$$

$$+(\underset{(0.00152)}{0.00868} - \underset{(0.00114)}{0.00458} * \mathrm{LEGSTAT} - \underset{(0.00173)}{0.00819} * \mathrm{FFP} * \mathrm{LEGSTAT}) * \mathrm{SRSHRS}$$

$$+(\underset{(0.0199)}{0.156} - \underset{(0.0478)}{0.141} * \mathrm{LEGSTAT} + \underset{(0.0555)}{0.398} * \mathrm{FFP} * \mathrm{LEGSTAT}) * \mathrm{VOLSTAF}$$

$$+ \underset{(0.0298)}{0.0388} * \mathrm{LEGSTAT} * \mathrm{DCVISIT} + \underset{(0.147)}{0.724} * \mathrm{SUMDUM}$$

$R^2 = 0.764 \quad F(10,728)=236$

SER/LHS MEAN=1.83/5.30=0.345

Where:

DCVISIT = Number of licensing/monitoring inspections in the past year.

Three of the four components of total cost are given in Equations (12) to (14). Monthly rental expenditures should vary directly with capacity and the number of groups required by federal group size regulations. Both variables imply a space requirement, whether legal as for licensed capacty, or practical. Monthly rental costs per unit of space rise with income as long as space measures such as CAP are included. State median family income was used as a regional index of economic conditions. The number of rental units of $100 per month or less in the zip code area in Equation (12) represents conditions in local rental markets for housing. With median family income in the zip code area absorbing variations for income differences, the number of rental units reflect population size as well. Rental expenses rise with the number of groups required to comply with the FIDCR group size limits. The increment is about $81 per month per group. Because this estimate is made holding licensed capacity (and hence rented space) constant, the increase must come from other sources. One possibility is that quality improvements in center facilities take place as centers grow. Another is that the need for specialized space (e.g., kitchen and bathroom facilities) rises with the number of groups.

For each additional year of operation, a center's monthly rent falls by $11. This may happen for several reasons. First, most older centers occupy older facilities and these generally command lower rents. Second, longer terms of occupancy, especially of leased facilities, usually cause fewer rent increases since the turnover point is often the time when rents can be raised with the least resistance. Third, older centers may own their own facilities in which case they may have responded to the survey question with an imputed rental cost below market. Finally, older centers staffed with more experienced administrators may simply show more business accumen in the choice of different facilities.

$$
\begin{aligned}
12.\quad \text{MONRENT} = {} & \underset{(171.)}{-419.} + \underset{(0.971)}{7.38*\text{CAP}} + \underset{(0.0182)}{0.0576*\text{MEDY}} + \underset{(21.8)}{81.3*\text{REQGRP}} \\
& + \underset{(.00624)}{0.0196*\text{FTCOST}} - \underset{(55.5)}{309.*\text{DONSPACE}} - \underset{(56.7)}{176.*\text{CLS2}} \\
& \underset{(58.2)}{-156.*\text{CLS4}} + \underset{(0.0123)}{0.0453*\text{RNT100A}} - \underset{(3.27)}{11.0*\text{YRSOPEN}} \\
& \underset{(68.9)}{-170.*\text{REG6}} - \underset{(73.6)}{185.*\text{REG9}}
\end{aligned}
$$

$R^2 = 0.39 \quad F(11,727)=42.3$

SER/LHS MEAN=571./499.=1.14

13. $$\text{MONSUP} = \underset{(186.)}{-656.} + \underset{(1.06)}{4.31}\ast\text{CAP} + \underset{(0.637)}{2.96}\ast\text{NMEALS} + \underset{(57.9)}{104.}\ast\text{SUMDUM}$$

$$+ \underset{(0.0191)}{0.0523}\ast\text{MEDY} + \underset{(11.3)}{79.8}\ast\text{CFTE}$$

$R^2 = 0.405$ F(5,733)=100.0 SER/LHS MEAN=727./760.=0.957

14. $$\text{OTHCOST} = \underset{(60.3)}{-113.} + \underset{(1.06)}{3.59}\ast\text{KFTE} + \underset{(78.4)}{270.}\ast\text{PART} - \underset{(57.3)}{96.6}\ast\text{SUMDUM} + \underset{(58.3)}{168}\ast\text{LEGSTAT}$$

$$+ \underset{(9.6)}{32.0}\ast\text{NCARE} + \underset{(2.58)}{9.47}\ast\text{TRANP} + \underset{(0.169)}{0.461}\ast\text{FFP}\ast\text{FRSHRS}$$

$R^2 = 0.147$ F(7,731)=18.0 SER/LHS MEAN=717./280.=2.56

Where:

DONSPACE = A dummy variable, equal to 1 if the center receives donated space, and equal to 0 otherwise.

CLS2 = A dummy variable, equal to 1 if the center is non-profit and non-FFP, and equal to 0 otherwise.

CLS4 = A dummy variable, equal to 1 if the center is non-profit, non-waiverable and FFP, and equal to 0 otherwise.

RNT100A = Number of housing units in center catchment area which rented for $100 per month in 1970.

YRSOPEN = Number of years the center has been operating.

REG6 = A dummy variable, equal to 1 if the center lies within HEW's Region 6, and equal to 0 otherwise.

REG9 = A dummy variable, equal to 1 if the center lies within HEW's Region 9, and equal to 0 otherwise.

Monthly supply expenditures, MONSUP, are explained primarily by the number of meals served, which can be the largest supply expense for smaller centers. Previous specifications of this equation included either ENR or KFTE as scale measures but each was insignificant and KFTE lowered the t-statistic for NMEALS below 1.5. An apparently strong correlation of NMEALS and KFTE makes the number of meals a strong instrumental variable for center size in addition to its importance for cost determination by itself. The number of caregivers approximates the number of classrooms, each of which has a supply requirements. Because many supplies such as toys cannot be shared across groups, this expense rises by $30 per month for each additional group.

Taxes, interest, insurance, fuel, and other operating expenses not included elsewhere are measured by OTHCOST. This residual category, of course, also includes measurement errors of the other components of cost. Miscellaneous costs depend heavily upon the legal status of a center because non-profit centers pay no taxes. Enrollment enters the equation primarily as a scale measure because most elements of OTHCOST are fixed and insensitive to short-run variations in enrollment. Parent participation in staff selection raises other costs by $270 per month, a surprising result. A negative effect on costs was anticipated here reflecting parental desire for less expensive operations. This strong positive influence may actually represent the cost differences of a class of centers not identified in the sample.

CHAPTER FOUR: MODEL SIMULATION

The essential first step in the simulation process, estimation of the model, has been presented as if carried out in isolation. But the structural coefficients actually play the key role in simulation by setting parameters for the equation system, thereby establishing the quantitative framework of industry behavior. This chapter describes the procedures and results for ten policy experiments. Each experiment represents one alternative to the existing federal regulations. These were selected as the most likely candidates for future regulatory changes, but by no means do they reflect the full array of possibilities which can be evaluated with this model. In fact, hundreds of reasonable configurations can be considered within the present model and very minor adaptations permit evaluation of many more. Some general modifications are discussed in Chapter Five.

Methodology

Simulating a policy change, in the simplest terms, requires that an algebraic solution be found for an equation at two or more values of a policy parameter. The process becomes more complex when additional equations are involved in a simultaneous system as in the present model. It is then necessary to find a vector containing only elements that satisfy the complete system simultaneously. As with the simple case, this vector must be computed for each policy experiment.*

---

*An important advantage of cross-sectional simulation is that a solution vector for one case does not depend upon the existence or quality of a solution in other cases. The presence of lagged endogenous variables in equations of a time-series model allows simulation error to accumulate. Divergences then occur which prohibit simulation of later periods without some initialization (resetting of starting values) taking place.

301

Before changes are made to policy parameters, actual data are applied to the unchanged structure to obtain "baseline" value for the endogenous variables, i.e., a simulation of the actual values of these variables. Ideally these baseline data duplicate the actual values used as input but since the system contains stochastic equations (that is, statements which include error terms), some divergence between actual and baseline (input and output) data is a certainty. The baseline data serve two functions. First, they provide a measure of the quality of the model by evaluating these divergences. Second, the output of this simulation will be the basis for evaluating policy impacts in the set of experiments. All measures of change in endogenous variables as the result of a policy change will be made with respect to the baseline.

A number of baseline simulations have been conducted, and each led to some adjustment of the original model specification. In particular, it was as a result of unsatisfactory simulated values for FRNO, ENR, and KFTE that GOVKIDS was recast as a function of predetermined variables only, effectively removing it from the large block containing most other endogenous variables. Generally, changes to the original model that were necessary to improve overall simulation quality led to poorer statistical fitting of the individual equations involved. This frequently occurs in simultaneous models and emphasizes the need to account for interactions among endogenous variables to fully explain the system being modeled.

The simulations were carried out in two rounds and only data for FFP centers were included. Table 2 describes completely the policy experiments carried out in each round. All simulations differ with respect to one, two, or three factors--child-to-staff ratios for target ages, group size limits for target ages, and/or one of three "enforcement levels".

Table 2

POLICY PARAMETER VALUES AND ENFORCEMENT LEVELS BY SIMULATION

ROUND I

Parameter Value		Simulation Baseline	Simulation I	Simulation II	Simulation III	Simulation IV
	Under 3	4:1	4:1	4:1	4:1	4:1
Child/	3 Years	5:1	7.5:1	5:1	7.5:1	7.5:1
Staff	4 Years	7:1	7.5:1	7:1	7.5:1	7.5:1
Ratio	5 Years	7:1	7.5:1	7:1	7.5:1	7.5:1
By Age	Over 5	20:1	20:1	20:1	20:1	20:1
	Under 3	15	15	15	15	15
Group	3 Years	15	15	15	15	15
Size	4 Years	20	20	20	20	20
Limit	5 Years	20	20	20	20	20
By Age	Over 5	25	25	25	25	25
Enforcement* Level		A	A	A	B	C

ROUND II

Parameter Value		Simulation Baseline	Simulation V	Simulation VI	Simulation VII	Simulation VIII	Simulation IX	Simulation X
	Under 3	4:1	4:1	4:1	4:1	4:1	4:1	4:1
Child/	3 Years	5:1	7:1	7:1	7:1	9:1	9:1	9:1
Staff	4 Years	7:1	7:1	7:1	7:1	9:1	9:1	9:1
Ratio	5 Years	7:1	7:1	7:1	7:1	9:1	9:1	9:1
By Age	Over 5	20:1	20:1	20:1	20:1	20:1	20:1	20:1
	Under 3	15	14	14	14	15	15	15
Group	3 Years	15	14	14	14	18	18	18
Size	4 Years	20	14	14	14	18	18	18
Limit	5 Years	20	14	14	14	18	18	18
By Age	Over 5	25	25	25	25	25	25	25
Enforcement* Level		A	A	B	C	A	B	C

*Enforcement level is one of three regimes:
A = Enforcement by current means and at current levels
B = All centers fallin below compliance with ratio and group size limits are forced to comply.
C = All centers whether under-or over-achieving compliance forced to exact compliance with regimes.

"Enforcement level" is one of three regimes:

A = Enforcement by current means and at current levels.

B = All centers falling below compliance with ratio and group size limits are forced to comply.

C = All centers whether under- or over-achieving compliance forced to exact compliance with regulations.

Round I shows output for profit and non-profit FFP centers separately to illustrate the important differences between the two types of operations and the remarkably predictable economic behavior of the profitmaking group. Round II combines the two categories to show how a typical center in the regulated portion of the market responds to policy changes.

Baseline Evaluation

A comparison of the baseline simulation and actual data is given in Table 3 for FFP profit and non-profit centers combined. Baseline values shown here are not comparable with those in subsequent tables because of differing samples. The ratio of the root mean square error (RMSE) to the standard error of the regression (SER) provides a simple and reliable test of the model's accuracy as a replicator of actual values. A value for this ratio of 2.0 or less will be considered acceptable. As can be observed even without the benefit of the ratio test, the model is able to predict the actual data of most variables with high precision. This high quality can be expected in the simulation of proposed policy changes as well.

The prediction for KFTE is just outside the acceptable limit and variables which depend upon it may suffer some loss in predictive quality. OTHCOST, a variable relying upon KFTE in a stochastic equation, has a ratio value

Table 3

COMPARISON OF ACTUAL DATA AND BASELINE SIMULATION DATA

Variable	Actual Value	Baseline Value**	Error Actual Minus Baseline	Root Mean Square Error	Standard Error of the Regression	RMSE/SER
ACSR	6.23	5.85	0.53	2.58	*	*
AMFR	0.31	0.24	–0.01	2.81	*	*
AMSR	2.94	3.00	–0.03	2.59	*	*
AVESAL1	510.81	503.96	5.09	117.34	106.00	1.11
AVSIZE	15.14	14.72	0.60	8.36	5.84	1.43
CFTE	6.29	6.50	–0.07	3.24	1.83	1.77
COSTPK	135.41	132.07	0.84	55.00	*	*
DEX	0.38	0.38	0.00	0.46	0.39	1.21
EDMEAN	13.54	13.57	–0.05	1.15	1.20	0.96
ENR	47.20	48.64	0.13	17.64	*	*
EXMEAN	3.86	3.89	0.01	2.48	3.11	0.80
FCSR	6.20	5.98	0.23	1.39	*	*
FOK	0.61	0.64	–0.05	0.56	*	*
FRNO	31.16	32.30	–0.03	21.72	18.50	1.17
FRSHRS	271.67	284.40	–3.32	120.60	*	*
FTEPDSTF	6.82	7.06	–0.07	3.25	1.84	1.77
GOVKIDS	16.04	16.34	0.16	14.63	8.96	1.63
HRSOP	10.71	10.73	0.00	0.87	0.92	0.95
KFTE	41.93	43.62	–0.23	17.13	8.53	2.01
KH3	443.98	471.78	-14.26	308.40	231.00	1.34
KH4	465.53	451.56	28.64	251.32	*	*
KH5	380.14	392.84	0.60	301.09	224.00	1.34
MONCOST1	5231.35	5485.81	-113.68	2569.20	*	*
MONRENT	483.16	539.01	-42.73	521.20	571.00	0.91
MONSUP	878.02	897.23	6.54	760.48	727.00	1.05
NCARE	2.72	2.84	–0.05	3.28	2.48	1.32
NMEALS	86.21	91.22	-1.94	42.80	46.60	0.92
NURSE	0.16	0.18	–0.01	0.37	0.36	1.03
OTHCOST	395.00	429.11	-16.98	1004.65	717.00	1.40
OVERS	172.80	188.69	-7.21	282.30	264.00	1.07
PARSERV3	0.63	0.63	0.00	0.48	0.45	1.07
PART	0.28	0.28	–0.01	0.42	0.33	1.27
PEX	0.36	0.37	–0.02	0.49	0.40	1.23
POORB	42.75	40.44	2.31	32.83	19.10	1.72
REQGRP	2.33	2.43	–0.02	0.97	*	*
REVIEW	0.44	0.44	–0.01	0.47	0.41	1.15
RICHB	16.50	17.65	–0.89	21.83	22.30	0.98
SALEXP	3475.33	3620.57	-60.51	1787.99	*	*
SCSR	11.52	11.06	0.57	2.18	*	*
SOK	0.93	0.98	–0.05	0.28	*	*
SRSHRS	152.30	158.58	-1.76	72.06	48.00	1.50
SW	0.18	0.21	–0.02	0.38	0.34	1.12
TRANP	6.03	6.80	–0.54	0.73	0.19	3.86
UNDER3	214.84	239.91	-17.20	338.51	309.00	1.10
USDAA	0.47	0.46	0.02	0.39	0.34	1.15
VOLSTAF	2.95	2.96	–0.04	4.76	4.03	1.18

---

* SER is not given for non-stochastic equations.

**Means computed on some cases omitted from simulation results.

of 1.4, however, which may indicate that the problem is not very serious. The predicted number of children to be transported, TRANP, is not reliable, but its linkages to other equations are not strong and so of little influence on the quality of other predictions.

Simulation Results

Comparison of the baseline to Simulations I to IV are given for profitmaking FFP centers in Table 4. The same information for non-profit FFP centers appears in Table 4. Table 5 compares the baseline with Simulations V to X for all FFP centers combined.*

- Profitmaking FFP Centers (Table 4) Round I

Considering profitmaking FFP centers first, Simulation I for the relaxed FIDCR ration requirement shows an average decline in required staff hours (FRSHRS), of 40 hours or one FTE caregiver as the initial impact. Centers respond by reducing actual caregiving staff by 0.55 of a person or 22 hours. Average monthly cost per child (COSTPK) has fallen by $6.69, and this is due mostly to the decline in salaries (SALEXP) and supply expenses (MONSUP). SALEXP drops almost exclusively because of a decline in the number of paid staff, average wages having fallen by only $1.50 per month.

Average actual child/staff ratio rises by 0.90 of an FTE child. Each of the five age groups contributes by rising about six hours per month while CFTE drops by more than one-half of a caregiver. The impacts on many variables

---

*Values for ACSR, SCSR, FCSR, COSTPK, and SALEXP cannot be computed from Tables 3 to 6 using the identities given in Appendix B. The values given in the tables for these variables and their components are mean values, and products or quotients of means do not necessarily equal the means of products or quotients.

Table 4

ROUND I FFP PROFIT CENTERS

Variable	Baseline	Simulation I	Simulation II	Simulation III	Simulation IV
ACSR	7.61	8.51	7.60	7.04	7.04
AMFR	-1.42	-1.09	-1.41	0.00	0.00
AMSR	2.17	1.57	2.17	2.59	2.58
AVESAL1	476.59	475.09	475.54	476.94	476.56
AVSIZE	15.72	15.79	15.40	14.96	15.35
CFTE	6.03	5.48	6.02	6.42	6.41
COSTPK	101.93	95.24	103.49	105.86	107.33
DEX	0.23	0.23	0.23	0.23	0.23
EDMEAN	13.51	13.50	13.51	13.51	13.51
ENR	58.18	58.91	58.01	57.75	57.49
EXMEAN	4.60	4.62	4.60	4.59	4.59
FCSR	6.12	7.04	6.11	7.04	7.04
FOK	0.00	0.02	0.00	1.00	1.00
FRNO	47.69	48.42	47.53	47.27	47.11
FRSHRS	338.72	298.55	337.86	291.85	291.16
FTEPDSTF	6.55	6.11	6.63	6.87	6.94
GOVKIDS	10.48	10.48	10.48	10.48	10.48
HRSOP	11.14	11.18	11.14	11.12	11.12
KFTE	52.13	52.88	51.99	51.68	51.55
KH3	526.38	532.75	525.16	522.57	521.44
KH4	501.80	506.83	500.77	498.53	497.56
KH5	469.65	476.36	468.22	465.78	464.44
MONCOST1	5,362.32	5,092.52	5,426.82	5,519.64	5,576.48
MONRENT	797.44	800.77	829.64	795.43	827.44
MONSUP	932.12	888.24	931.65	963.63	962.45
NCARE	2.12	2.16	2.12	2.11	2.10
NMEALS	100.83	100.77	100.85	100.86	100.88
NURSE	0.13	0.13	0.13	0.13	0.13
OTHCOST	494.76	482.82	493.71	472.92	471.61
OVERS	260.95	266.85	259.95	257.40	256.48
PARSERV3	0.51	0.50	0.51	0.51	0.51
PART	0.03	0.03	0.03	0.03	0.03
PEX	0.21	0.21	0.21	0.21	0.21
POORB	27.34	27.09	27.40	27.44	27.50
REQGRP	2.90	2.94	3.29	2.87	3.27
REVIEW	0.19	0.19	0.19	0.19	0.19
RICHB	23.56	23.67	23.54	23.52	23.49
SALEXP	3,138.00	2,920.69	3,171.84	3,287.66	3,314.99
SCSR	11.95	11.99	11.95	11.93	11.93
SOK	0.97	0.91	0.97	1.00	1.00
SRSHRS	175.48	177.61	175.08	174.06	173.68
SW	0.14	0.14	0.14	0.14	0.14
TRANP	6.53	6.61	6.52	6.50	6.48
UNDER3	326.49	332.34	325.63	322.93	322.16
USDAA	0.00	0.00	0.00	0.00	0.00
VOLSTAF	1.81	1.87	1.82	1.82	1.79

are negligible. Service variables do not change except for a slight decline in meals served and a rise in the number of chldren transported. Staffing characteristics (EDMEAN, EXMEAN) and average group size (AVSIZE) are unaffected.

The impacts of smaller maximum group size, Simulation II, are transmitted to a minor increase in COSTPK ($1.56) in two steps: the number of required center groups (REQCRP) rises by 0.39, and rental cost (MONRENT) rises as space expands to accommodate the added groups. Some expansion of paid caregiving staff occurs as well, and average size of groups containing three- and four-year olds declines slightly. The effects of this policy change are very small for two possible reasons. First, profitmaking centers that would comply with the new regulation may have some excess space in affected groups. Second, the change in AVSIZE shows that, on average, these centers will move toward but not achieve compliance.

Simulation III gives the impact of eliminating non-compliance with both ratio and group size regulations in those centers where it exists. Recall that ratio requirements are relaxed as in Simulation I, and group sizes are set by the schedule used in Simulation II. The results are striking for their clarity. Average group size (AVSIZE) is forced below 15 children for the target age groups. The difference between actual and federal required staff is exactly zero. Characteristically, profitmaking centers would do no more than is absolutely required to satisfy these regulations. Cost per child rises by nearly $5 per month and this increase is caused primarily by a rise in caregiving staff which raises payroll and supply expenses.

The drop in the value of the policy variable FRSHRS is purely a second round effect of the enforced compliance with ratio requirements. A small decline in the hours of attendance of each age group occurs as centers reduce total enrollment somewhat. Simultaneously, the size of their caregiving staffs rises to complete the adjustment.

The last experiment for profitmaking centers simulates exact compliance with ratio requirements of Simulation I and group size limits of Simulation II. Cost per child is highest for this policy with the sources of the increase similar to those in Simulation III. Centers previously in overcompliance with one or both aspects of the regulations (ratio and group size) seem to be unable to restore average group size to less than 15 children when forced to discard excess staff. It is the formerly overcomplying centers creating the problem since nothing has altered the conditions of the noncompliers from Simulation III.

- Non-Profit FFP Centers (Table 5) Round I

The same sort of substantial impacts of relaxed ratio requirements found among profitmaking centers emerge in Simulation I for non-profit centers. The decline in salary expenses is the greatest source of reduced average cost and is primarily due to a drop in the number of caregivers. Supply expenses also fall, chiefly because the decline in caregivers reduces the demand for classroom supplies. Again, services and staff characteristics are unchanged.

The response patterns in Simulations I and II will follow very closely the patterns of the profitmaking centers. This is so because of model design in some clases--coefficients of certain terms are the same for both center types. In other instances, behavior of centers is not determined by their legal status but by practical necessity. For example, any center will usually accept an opportunity to cut costs whether it is a profit-maximizer or not. Simple mechanics of center operation take over at that point and these may be qualitatively similar in the two types of centers. Quantitative differences in impacts are almost a certainty, however, and in some instances they may be sizeable.

Table 5

FFP NON-PROFIT CENTERS

Variable	Baseline	Simulation I	Simulation II	Simulation III	Simulation IV
ACSR	5.38	5.89	5.38	5.86	7.15
AMFR	0.76	1.08	0.76	1.12	0.00
AMSR	3.70	3.12	3.69	3.16	2.12
AVESAL1	523.70	522.16	522.85	522.15	518.32
AVSIZE	15.06	15.11	14.72	14.52	14.87
CFTE	7.48	6.96	7.46	7.00	6.08
COSTPK	139.06	130.21	140.84	130.51	114.24
DEX	0.44	0.44	0.44	0.44	0.44
EDMEAN	13.67	13.67	13.67	13.67	13.65
ENR	53.07	54.04	52.88	54.00	55.78
EXMEAN	3.45	3.47	3.45	3.47	3.50
FCSR	6.13	7.14	6.13	7.14	7.15
FOK	0.88	0.95	0.88	1.00	1.00
FRNO	33.63	34.59	33.43	34.56	36.34
FRSHRS	305.30	267.21	304.33	267.02	276.17
FTEPDSTF	8.10	7.68	8.18	7.71	7.07
GOVKIDS	19.44	19.44	19.44	19.44	19.44
HRSOP	10.58	10.60	10.58	10.60	10.65
KFTE	47.15	48.01	46.99	47.98	49.63
KH3	511.25	518.76	509.88	518.48	532.77
KH4	492.55	498.67	491.39	498.38	509.79
KH5	428.01	436.46	426.38	436.18	451.97
MONCOST1	6,302.29	6,020.66	6,360.41	6,037.54	5,652.18
MONRENT	562.78	566.61	593.16	566.46	606.36
MONSUP	1,036.28	944.38	1,034.67	977.21	923.16
NCARE	3.42	3.43	3.42	3.43	3.45
NMEALS	99.14	99.02	99.16	99.02	98.79
NURSE	0.20	0.20	0.20	0.20	0.20
OTHCOST	446.22	433.85	444.75	433.61	446.34
OVERS	210.00	216.51	208.89	216.25	228.88
PARSERV3	0.67	0.66	0.67	0.66	0.65
PART	0.38	0.38	0.38	0.38	0.38
PEX	0.41	0.41	0.41	0.41	0.41
POORB	44.88	44.43	44.98	44.43	43.33
REQGRP	2.62	2.67	2.99	2.67	3.16
REVIEW	0.55	0.55	0.55	0.55	0.55
RICHB	16.10	16.33	16.06	16.33	16.89
SALEXP	4,257.02	4,025.83	4,287.84	4,040.27	3,676.32
SCSR	11.29	11.34	11.28	11.34	11.46
SOK	1.00	1.00	1.00	1.00	0.93
SRSHRS	171.40	174.05	170.91	173.98	179.41
SW	0.23	0.23	0.23	0.23	0.22
TRANP	7.99	8.11	7.96	8.11	8.31
UNDER3	241.15	250.14	243.20	249.89	261.81
USDAA	0.66	0.66	0.66	0.66	0.66
VOLSTAF	3.65	3.65	3.63	3.65	3.68

Simulation II shows AVSIZE dropping below 15 children in clear reaction to stricter grouping limits. Unlike the profitmakers, the average non-profit center is able to comply with the tighter requirement. No other important changes take place, again, because these centers may presently possess the capacity for more restrictive group limits without substantial increases in staff or space.

Non-profit reaction to forced compliance (Simulation III) is remarkably close to the response to relaxed ratio requirements. This happens because the effect of the Simulation I policy is to raise the compliance rate, FOK, to 95 percent while the stimulus of the Simulation III policy is to force the compliance rate to 100 percent. The causes of the changes in equilibria differ but the new equilibria are nearly identical.* This suggests that policy designs which differ in fundamental ways might be used interchangeably when one or the other is politically or economically infeasible.

Simulation IV gives clear indication of how far above simple compliance with ratio and group size the non-profit centers actually are. The best evidence of this is the sharp drop in cost per child--on average, a $25 reduction per month. This causes an increase in enrollment of parent-fee children in all age groups which, incidentally, causes another round of decreases in average cost.** The increased enrollment forces the creation of new groups as required by the regulations.

---

*Simulation III includes the stricter group size limits which have already been shown to be inconsequential. The five percentage point difference in compliance rates between the policies is apparently too small to have an effect of its own.

**There may be very many rounds of enrollment-cost-enrollment cycles, each one slightly smaller than the last until the changes are infinitesimal and equilibrium is reached.

Both lower salary and smaller supply expenditures are the result of cutbacks in caregiver staff. Rental costs increase on the average, responding to the larger number of required groups. The net change in total cost is an average decline of $650 per month. This reinforces the reduction in COSTPK caused by increased enrollment. The entire set of impulses and reactions cease when exact compliance with the two regulatory measure is achieved. Note, for example, that ACSR = FCSR, meaning that child-to-staff ratio has stabilized at the required level. Group size compliance is not as easily demonstrated but AVSIZE now has a value below and very near 15, showing that for three- and four-year olds, at least, the requirement is being met.

- All FFP Centers (Table 6) Round II

Simulations V, VI, and VII involve a minor relaxation of the ratio requirement for three-year olds, and substantial tightening of group size limits for four- and five-year olds. When current enforcement practices are used, average cost per child drops by $3.65 per month or about three percent, chiefly due to a decline in payroll and supply expenses. For enforcement level B, the cost decrease is much smaller, less than one percent, because the savings found in V are offset by undercomplying centers' efforts to satisfy the regulations. The changes in ACSR and AVSIZE from V to VI give the magnitudes of the adjustments in ratio and group size respectively. ACSR falls by 5.3 percent and AVSIZE by 8 percent indicating that significant undercompliance with both regulations was present.

Simulation VII shows the effects of eliminating over- as well as undercompliance. Monthly average cost falls 10.7 percent from the baseline value and 10 percent from the regime B level. This policy also brings about a clear though small shift in the income distribution of enrollment. The percent of children coming from families with incomes less

Table 6

ROUND II - ALL FFP CENTERS

	FFP Center Baseline	Simulation V	Simulation VI	Simulation VII	Simulation VIII	Simulation IX	Simulation X
ACSR	5.84	6.23	5.90	6.78	6.96	6.56	8.09
AMFR	0.30	0.52	0.81	0.00	0.87	1.07	0.00
AMSR	3.40	3.00	3.27	2.52	2.37	2.53	1.55
AVESAL1	513.75	511.29	510.93	510.04	510.86	513.39	512.06
AVSIZE	15.15	14.73	13.52	14.00	15.27	14.62	18.00
CFTE	7.24	6.86	7.11	6.46	6.31	6.38	5.58
COSTPK	131.19	127.54	130.08	117.11	116.15	128.08	103.12
DEX	0.40	0.39	0.40	0.39	0.39	0.40	0.39
EDMEAN	13.61	13.61	13.61	13.60	13.60	13.60	13.59
ENR	54.70	55.10	54.82	56.23	56.34	55.04	57.76
EXMEAN	3.66	3.67	3.66	3.68	3.69	3.69	3.71
FCSR	6.13	6.77	6.77	6.78	8.09	8.06	8.09
FOK	0.69	0.73	1.00	1.00	0.78	1.00	1.00
FMNO	37.06	37.46	37.18	38.60	38.70	37.40	40.12
FRSHRS	315.35	287.97	286.32	293.75	246.98	241.44	253.65
FTEPOSTF	7.84	7.67	7.87	7.36	7.09	7.16	6.52
GOVKIDS	17.64	17.64	17.64	17.64	17.64	17.64	17.64
HRSOP	10.69	10.71	10.70	10.73	10.74	10.72	10.78
KFTE	48.71	49.10	48.81	50.10	50.22	49.09	51.56
KH3	518.55	521.89	519.47	530.56	531.65	521.91	543.21
KH4	497.90	500.56	498.49	507.49	508.51	499.30	517.58
KH5	440.97	444.55	442.06	454.55	455.56	444.69	468.20
MONCOST1	6,159.20	6,058.41	6,165.88	5,856.13	5,662.11	5,700.90	5,319.61
MONRENT	619.90	667.82	666.33	673.10	625.50	620.49	621.43
MONSUP	1,024.89	994.44	1,014.80	962.46	950.03	957.42	891.61
NCARE	3.18	3.17	3.16	3.18	3.20	3.24	3.25
NMEALS	100.02	99.97	99.98	99.80	99.83	100.21	99.65
NURSE	0.18	0.18	0.18	0.18	0.18	0.18	0.18
OTHCOST	460.71	450.52	448.18	458.50	439.10	433.88	450.34
OVERS	226.17	229.19	226.96	236.74	237.69	229.48	247.98
PARSERV3	0.64	0.63	0.63	0.63	0.62	0.63	0.62
PART	0.32	0.32	0.32	0.32	0.32	0.32	0.32
PEX	0.37	0.37	0.37	0.37	0.37	0.38	0.37
POORB	41.29	41.02	41.10	40.28	40.46	42.19	39.71
REQGRP	2.70	3.30	3.28	3.36	2.77	2.71	2.85
REVIEW	0.48	0.48	0.48	0.47	0.47	0.48	0.48
RICHB	17.71	17.84	17.80	18.22	18.12	17.25	18.50
SALEXP	4,053.79	3,945.63	4,036.58	3,762.09	3,647.48	3,689.11	3,346.23
SCSR	11.42	11.45	11.43	11.51	11.51	11.36	11.60
SOK	0.99	0.98	1.00	0.95	0.98	0.99	0.91
SRSHRS	174.05	175.21	174.38	178.67	178.61	175.10	183.01
SW	0.21	0.21	0.21	0.21	0.21	0.21	0.21
TRANP	7.82	7.87	7.84	8.00	8.03	7.90	8.19
UNDER3	264.85	267.77	265.57	274.69	275.66	268.34	285.45
USDAA	0.53	0.53	0.53	0.52	0.52	0.53	0.52

than $6,000 falls from 41.3 percent to 40.3 percent while the percent from high income homes (over $15,000 annual income) rises from 17.7 to 18.2 percent. Middle-income children increase by one-half of one percentage point as well. This shift is a result of growing private enrollment rather than of declines in the enrollment of poor children. The model constrains enrollment of government-funded children to be constant under the policy alternatives examined here, therefore, these changes in income mix may not be observed where GOVKIDS is subject to fluctuations in costs or private enrollment.

The increase in private enrollment (FRNO) is to be expected when costs fall since fees would drop as well in this case, and demand from households would rise in response.

Simulations VII, IX, and X relax the ratio requirements for ages three- to five-year olds to nine to one. Group size limits are also relaxed for these ages. Cost per child drops 11.5 percent from baseline, chiefly because of a sizeable increase in the actual child-to-staff ratio. AVSIZE rises very slightly in response to the higher group size limits. The increase in ACSR and drop in costs is accomplished primarily by a reduction in caregiving staff by nearly one full-time person while FTE enrollment rises by 1.5 children.

Enforcement level B applied to this policy leads to only a moderate decline in average cost from baseline tnad the savings is actually $12 per month less than that of VIII. Centers' apparent response to forced compliance with this policy is to cut FTE enrollment by one person from the enrollment level in Simulation VIII. No great change in other decision variables takes place when enforcement is tightened. A small change in CFTE leaves payroll nearly unchanged, and average group sizes for target ages change more as a result of falling enrollment than because the number of groups has increased.

The most substantial cost savings is obtained by enforcing exact compliance, regime C, with this policy. Monthly average cost is cut by more than 20 percent from baseline and by nearly this much over the same policy enforced for undercompliers only. This latter comparison gives an indication of the extent and costs of overcompliance. The difference in average costs between regimes B and C, about $25 per FTE child, originates from reductions in personnel and costs related directly to them. Some increase in FTE enrollment contributes to this drop but this is primarily a second round response to initial cost-cutting brought about by reducted payroll expenses. The $25 difference represents the average cost of overcompliance per child among FFP centers.

CHAPTER FIVE: CONCLUSION

The clearest outcomes of the Round I simulation experiments are the comparatively large impacts of altered child/staff ratio requirements and the nearly complete lach of reaction to tightened group size limits. These results show up with equal force for both profit and non-profit centers and suggest three important points:

1. The differential impacts between profit and non-profit centers from either of these two policies are minimal--no major shifts of resources from one center type to the other should occur as a result of either policy;
2. Group size alterations are nearly costless within the range analyzed here and cause comparatively few changes in other center characteristics; and
3. Manipulation of child/staff ratios can serve as an effective cost-control measure over a reasonable range and can be used to make FFP centers more competitive with private centers by substantial easing of the requirement. Our results show that characteristics other than staffing and enrollment are not altered very much by this policy.

Qualitative differences in impacts between profit and non-profit centers are obvious for policies involving forced compliance with regulations. These results are to be expected based on the differing motives for center operation. Profitmaking centers, by and large, do not comply with existing FIDCR regulations and, on average, their costs rise sharply when forced to comply with them. Non-profit centers feel no cost effects of regulation enforcement and actually show a substantial drop in costs when over-complying centers are required not to exceed standards. The ramifications for the day care market are potentially serious if policies of enforced compliance are adopted. Profitmaking FFP centers will lose competitive ground to private centers as FFP centers' average costs rise and they also suffer loss of much of the competitive price advantage held over non-profit FFP centers.

The responses to changing market conditions are more difficult to identify in the aggregated Round II output but some patterns in average cost are apparent nonetheless. Regardless of the policy configuration of ratio and group size chosen, the savings realized by eliminating overcompliance greatly outweighs the costs incurred from eliminating all undercompliance. Clearly, overcompliance is far more common among FFP centers than is noncompliance, at least in terms of the total numbers of children and staff involved if not in terms of the number of centers.* While enforcement of exact compliance, regime C, is the least feasible practice, the substantial cost reductions may justify a policy that sought to approach this level. In fact, the savings which result from a policy of eliminating overcompliance only is greater than that shown in the two regime C experiments. To the $25 savings which emerges from movement from Simulation IX to X must be added the difference between VIII and IX costs since this is the cost of eliminating undercompliance--an amount which would not be incurred if this problem were ignored. This brings the average cost saving per child to $37 per month. A smaller amount, $15.51, would be realized from similar enforcement of the policy used in V, VI, and VII.

Of necessity, this report addresses only a very narrow band of the policy alternatives testable with the simulation model. A most interesting set of new experiments might involve changes in reimbursement rates and practices such as sliding fee arrangements. The variable FTCOST (statewide expenditures of federal dollars on day care) is a potentially powerful policy instrument which has not been tested at all in the present work. Center participation in the USDA food program (another dormant instrument here) provides additional opportunities for experimentation.

---

*A net surplus of 29,000 full-time equivalent caregivers nationwide in FFP centers exists with respect to current FIDCR regulations. This represents more than 25 percent of the entire caregiver workforce (Coelen, Glantz, Calore, 1978).

An econometric variation might be undertaken in future work which opens up a whole new source of information to the policymaker. By re-casting the data and equation specifications into a time series model, not only are the equilibrium values of all variables obtainable, but so too are the speeds with which these variables approach their new equilibria. This information becomes particularly valuable when implementation time is an important consideration in policy choice.

Finally, and undoubtedly of greatest importance, is the re-estimation of the structural coefficients using a two-stage, least-squares technique, a method that accounts for the simultaneous interactions that actually take place. We explain elsewhere the necessity and practicality of the simpler, less expensive OLS approach actually used. However, a study of this scope and importance would surely benefit from a thorough search for tne best available analytical tools.

APPENDIX A

GLOSSARY OF VARIABLES APPEARING IN MODEL EQUATIONS

ACSR:	Actual child/staff ratio.
AMFR:	Actual minus FIDCR-required staff members.
AMSR:	Actual minus state-required staff members.
AVESAL1:	Average monthly staff salary.
AVRSIZE:	Average group size for groups containing three- or four-year old children.
CAP:	Licensed capacity.
CFTE:	Full-time equivalent (FTE) caregivers.
CLS2:	A dummy variable, equal to 1 if the center is non-profit and non-FFP, and equal to 0 otherwise.
CLS3:	A dummy variable, equal to 1 if the center is profit and non-FFP, and equal to 0 otherwise.
CLS4:	A dummy variable, equal to 1 if the center is non-profit, non-waiverable, and FFP, and equal to 0 otherwise.
CLS5:	A dummy variable, equal to 1 if the center is profit, waiverable, and FFP, and equal to 0 otherwise.
CLS6:	A dummy variable, equal to 1 if the center is non-profit, waiverable, and FFP, and equal to 0 otherwise.
COLPCT:	Percent of state population with a college degree.
COSTPK:	Average monthly expenditure per FTE child.
DCVISIT:	Number of licensing/monitoring inspections in the past year.
DEX:	A dummy variable, equal to 1 if the center offers dental exams, and equal to 0 otherwise.
DGSZ:	A dummy variable, equal to 1 if average center group size is less than 40, and equal to 0 otherwise.
DK40:	A dummy variable, equal to 1 if center size is greater than 40, and equal to 0 otherwise.
DNEW:	A dummy variable equal to 1 if the center is less than 2 years old, equal to 0 otherwise.

DONSPACE:	A dummy variable equal to 1 if the center receives donated space, equal to 0 otherwise.
EDMEAN:	Average level of education of center caregivers.
ENR:	Headcount enrollment.
EXMEAN:	Average years of experience of center caregivers.
FAMB:	Number of families served
FCSR:	FIDCR-required child/staff ratio
FEEBAR:	State average parent fee per child.
FFP:	A dummy variable, equal to 1 if the center is an FFP center, and equal to 0 if it is not.
FGAP:	Shortage of FTE caregivers with respect to current FIDCR staff/child ratio requirement.
FOK:	A dummy variable equal to 1 if the center complies with current FIDCR child/staff ratio requirements, and 0 otherwise.
FRNO:	Enrollment paid for by private funds.
FRSHRS:	FIDCR-required staff hours.
FTCOST:	Total Title XX expenditures for full-time placements in center care.
FTEPDSTF:	FTE paid staff.
GOVKIDS:	Enrollment paid for by government funds.
HRSOP:	Number of hours per day the center is open.
KFTE:	FTE enrollment.
KH3:	Total number of hours scheduled per week for three-year olds.
KH4:	Total number of hours scheduled per week for four-year olds.
KH5:	Total number of hours scheduled per week for five-year olds.
LEGSTAT:	A dummy variable, equal to 1 if the center is a profit center, equal to 0 if it is not.
MEDY:	Median annual family income in center catchment area.
MONCOST1:	Total monthly center expenditures.

MONRENT:	Monthly center expenditure for rent.
MONSUP:	Monthly center expenditure for supplies.
NCARE:	Number of non-caregivers.
NMEALS:	Number of meals served to full-time children per week.
NURSE:	A dummy variable, equal to 1 if the center has a nurse on staff, and equal to 0 if it does not.
OTHCOST:	Monthly center expenditures for miscellaneous items.
OVER5:	Total number of hours scheduled per week for children over five years of age.
PART:	A dummy variable, equal to 1 if parents participate in staff selection, and equal to 0 otherwise.
PARSERV3:	A dummy variable, equal to 1 if the center offers assistance to parents in obtaining general financial aid, and equal to 0 otherwise.
PEX:	A dummy variable, equal to 1 if the center offers physical exams, and equal to 0 otherwise.
POORB:	Percent of families with incomes of $6,000 per year or less served by the center.
REG1:	A dummy variable, equal to 1 if the center lies within HEW's region 1, and equal to 0 otherwise.
REG2:	A dummy variable, equal to 1 if the center lies within HEW's region 2, and equal to 0 otherwise.
REG3:	A dummy variable, equal to 1 if the center lies within HEW's Region 3, and equal to 0 otherwise.
REG5:	A dummy variable, equal to 1 if the center lies within HEW's Region 5, and equal to 0 otherwise.
REG6:	A dummy variable, equal to 1 if the center lies within HEW's Region 6, and equal to 0 otherwise.
REG9:	A dummy variable, equal to 1 if the center lies within HEW's Region 9, and equal to 0 otherwise.
REQGRP:	Number of FIDCR-required class groups.
REVIEW:	A dummy variable, equal to 1 if parents review center programs and budgets, and equal to 0 otherwise.

RICHB: Percent of families served which earn over $15,000 per year.

RNT100A: Number of housing units in center catchment area which rented for $100 per month in 1970.

SALEXP: Monthly center payroll expenditures.

SCSR: State-required child/staff ratio.

SGAP: Shortage of FTE caregivers with respect to state staff/child ratio requirement.

SOK: A dummy variable, equal to 1 if center complies with state staff/child ratio requirements, and equal to 0 otherwise.

SRSHRS: State-required staff hours.

SR3: Effective state-required staff/child ratio for three-year olds.

SR4: Effective state-required staff/child ratio for four-year olds.

SR5: Effective state-required staff/child ratio for five-year olds.

STATE: State identification number.

SW: A dummy variable, equal to 1 if the center has a social worker on staff, and equal to 0 if it does not.

TRANP: Number of children for whom transportation is provided.

UNDER3: Total number of hours scheduled per week for children under three years old.

USDAA: A dummy variable, equal to 1 if the center participates in the USDA food program, and equal to 0 if it does not.

VOLSTAF: Number of volunteers.

WELFIX2: Weekly welfare fee per child.

YRSOPEN: Number of years the center has been operating.

ZIPDUM4: A dummy variable, equal to 1 if the center is located in a SMSA, and equal to 0 otherwise.

APPENDIX B

DATA AND METHODS FOR EQUATION ESTIMATION

The source for the sample is a data base of 3167 day care centers surveyed from April 1976 to March 1977 (Coelen, Glantz, and Calore, 1978). A subsample of 841 centers was culled from the original base, selected for completeness of data on a series of 21 critical variables and for economy of estimation. A second screening eliminated 102 centers with data missing for any one of 90 variables appearing in later versions of the econometric model.

The econometric model has 46 equations and endogenous variables and 34 predetermined variables. Thirty-two equations are stochastic. Among the stochastic equations are 13 that summarize the industry's structure by describing outputs, inputs, and prices. Appendix C shows the list of equations.

Stochastic equations were originally specified to test for differences in the structure of behavioral relationships across different center types, geographic regions, etc. Early specifications were of the general form:

$$Y = a_o + \sum_{j=1}^{k-1} a_j W_j + \sum_{j=1}^{m} [(b_{io} + \sum_{j=1}^{k-1} b_{ij} W_j) \cdot X_i]$$

Where: Y = dependent variable

W_j = dummy variable for center type, region, etc.,

X_j = explanatory variable endogenous or predetermined,

a_j = intercept parameter,

b_{ij} = slope parameter.

For example, equation 3 as initially specified had:

$k = 4$ (one of four center types, non-FFP profit, FFP non-profit, etc.)

$m = 4$ (FRSHRS, SRSHRS, VOLSTAF, DCVISIT).

A third dummy for the summer survey was added and insignificant coefficients were dropped following the first estimations. The final equation by center type is:

A. NON-FFP PROFIT:

$$\text{CFTE} = 0.798 + 0.00955 * \text{FRSHRS} + 0.00410 * \text{SRSHRS} + 0.015 * \text{VOLSTAF} + 0.0388 * \text{DCVISIT} + 0.724 * \text{SUMDUM}$$

B. NON-FFP NON-PROFIT:

$$\text{CFTE} = 0.798 + 0.00955 * \text{FRSHRS} + 0.00868 * \text{SRSHRS} + 0.156 * \text{VOLSTAF} + 0.724 * \text{SUMDUM}$$

C. FFP PROFIT:

$$\text{CFTE} = 0.798 + 0.01442 * \text{FRSHRS} + 0.00409 * \text{SRSHRS} + 0.413 * \text{VOLSTAF} + 0.388 * \text{DCVISIT} * 0.724 * \text{SUMDUM}$$

D. FFP NON-PROFIT:

$$\text{CFTE} = 0.798 + 0.01442 * \text{FRSHRS} + 0.00868 * \text{SRSHRS} + 0.015 * \text{VOLSTAF} + 0.724 * \text{SUMDUM}$$

Other equations such as that for AVESAL1 and the socioeconomic status equations for POORB and RICHB were similar initial forms to account for potential interactions.

Identities in the equation list are specified for one of three reasons. Equations like those for MONCOST1, COSTPK, and ENR are used to close the equation set and as such are straightforward algebraic statements for definition. KH4, also of this type, is computed as a residual of other age variables in Equation (25). It was chosen for the left-hand side of the identity over the other components because it takes on the largest values on average. Errors in the other components would have relatively less effect upon its value than on that of another smaller component used in its place. Other identities are used for metering of simulation results. They generally are not part of a feedback network. Finally, identities describing policy in terms of model variables serve as the input route for policy parameters. Equations for FRSHRS and REQGRP and the stochastic equation for SRSHRS can be manipulated to simulate a wide range of federal and state policy changes. Each statement has linkages back into the model by which other parts of the system receive the policy impacts.

A number of estimation problems and their effects upon the model deserve discussion at this point. Most important among these is the bias and inconsistency of the ordinary least squares estimators of structural parameters in simultaneous equation systems. The error terms of these equations will be correlated with endogenous variables appearing on the right-hand side of the same equations. Since the model contains a large block of simultaneous relations, the bias problem is expected to be widespread. Two-stage least squares or simple instrumental variables procedures are two methods which provide simultaneous equation estimators which are unbiased.

Several equations in the system will possess heteroskedastic error structures which render OLS estimators inefficient and variance estimates biased. These equations are principally in two groups: total cost components and service equations. In the first group, unequal error variances arise because large centers will have more widely fluctuating spending patterns than small centers. This is due to the simple magnitude and variability of the costs they sustain.

The service equations with binary dependent variables comprise the second group. Because the left-hand side of one of these expressions is either zero or one, the error term will take only two values given the vector of explanatory variables. For example:*

y = 1 if service is provided by center

y = 0 if service is not provided

x = explanatory variable,

ϵ = disturbance term,

ϵ_i = structural parameters,

$y = \beta_o + \beta_1 x + \epsilon$

when y = 1, $= 1 - (\beta_o + \beta_1 x)$

when y = 0, $= \; - (\beta_o + \beta_1 x)$

If $E(\epsilon) = 0$, the probability of the first value must be $\beta_o + \beta_1 X$ and of the second $1 - \beta_o - \beta_1 X$. But $\beta_o + \beta_1 X$ may have a value outside the 0,1 range because no restriction is placed on OLS parameter estimates to force this term to behave as a probability predictor. The resultant error distribution differs sharply according to the value of the dependent variable.

Several corrective approaches might be taken to remove the heteroskedasticity from the equation where it occurs. Weighted least squares weights each observation by the estimated error variance for the group from which it is drawn. In the case of cost components, observations on large centers would be weighted by the estimated variance of errors for this group and so forth. Logit regressions may be performed on the binary variable equations.

In spite of the theoretical inferiority of OLS estimation for the problems discussed above, the practical limitations of time and money make this method preferable to the 2SLS, GLS, and logit procedures in their respective contexts. In fact, in Monte Carlo studies of small sample properties of estimators discussed in Johnston,* these more sophisticated techniques have not proven to be superior to ordinary least squares when a variety of estimation problems exist. Ordinary least squares, then, is the method used for estimation of all parameters. The simulation process described in Chapter Four made iterative adjustments to the original model specification necessary in order to obtain simultaneous solutions to some equations. The adjustments, involving respecification and re-estimation in a number of cases, actually were an improvement upon the straight OLS procedure because constraints not imposed on coefficients by OLS were effectively imposed by the adjustments and subsequent simulation attempts. Nevertheless, caution should be used when evaluating the results of parameter estimation and hypothesis tests reported above.

---

*The example is found in Thiel, Principles of Econometrics, Wiley, New York, 1971.

---

*Johnston, J, Econometric Methods, McGraw Hill, New York, 1971.

APPENDIX C

EQUATIONS

1.
$$\begin{aligned}\mathrm{FRNO} = {} & \underset{(4.72)}{-9.27} + (\underset{(0.0227)}{0.687} + \underset{(0.0269)}{0.08} * \mathrm{LEGSTAT}) * (1-\mathrm{FFP}) * \mathrm{CAP} \\ & + \underset{(0.0274)}{0.7} * \mathrm{FFP} * (\mathrm{CAP} - \mathrm{GOVKIDS}) + \underset{(0.000519)}{0.0033} * \mathrm{MEDY} - \underset{(0.013)}{0.109} * \mathrm{COSTPK} \\ & - \underset{(1.52)}{3.48} * \mathrm{ZIPDUM4}\end{aligned}$$

$R^2 = 0.699$ F(6,732)=284. SER/LHS MEAN=18.5/40.5=0.457

2.
$$\begin{aligned}\mathrm{GOVKIDS} = {} & \mathrm{FFP} * \mathrm{CAP} * (\underset{(0.0339)}{0.00119} + \underset{(0.0277)}{0.185} * \mathrm{WELFIX2}/\mathrm{FEEBAR} \\ & + \underset{(0.96)}{7.15}/\mathrm{CAP} - \underset{(0.0179)}{0.137} * \mathrm{LEGSTAT} * \underset{(2.45)}{4.25} * \mathrm{FTCOST}/1000000)\end{aligned}$$

$R^2 = 0.522$ F(4,734)=200.0 SER/LHS MEAN=8.96/6.25=1.43

3.
$$\begin{aligned}\mathrm{CFTE} = {} & \underset{(0.139)}{0.798} + (\underset{(0.000814)}{0.00955} + \underset{(0.000539)}{0.00487} * \mathrm{FFP}) * \mathrm{FRSHRS} \\ & + (\underset{(0.00152)}{0.00868} - \underset{(0.00114)}{0.00458} * \mathrm{LEGSTAT} - \underset{(0.00173)}{0.00819} * \mathrm{FFP} * \mathrm{LEGSTAT}) * \mathrm{SRSHRS} \\ & + (\underset{(0.0199)}{0.156} - \underset{(0.0478)}{0.141} * \mathrm{LEGSTAT} + \underset{(0.0555)}{0.398} * \mathrm{FFP} * \mathrm{LEGSTAT}) * \mathrm{VOLSTAF} \\ & + \underset{(0.0298)}{0.0388} * \mathrm{LEGSTAT} * \mathrm{DCVISIT} + \underset{(0.147)}{0.724} * \mathrm{SUMDUM}\end{aligned}$$

$R^2 = 0.764$ F(10,728)=236. SER/LHS MEAN=1.83/5.30=0.345

NOTE: Standard errors in parentheses.

4. $$\text{NCARE} = \underset{(1.09)}{-4.03} + \underset{(0.1)}{0.325}*\text{HRSOP} + \underset{(0.00303)}{0.0277}*\text{FRNO} + \underset{(0.553)}{2.54}*\text{GOVKIDS/ENR}$$

$$+ \underset{(0.227)}{0.919}*\text{CLS2} + \underset{(0.337)}{0.982}*\text{CLS4} + \underset{(0.0222)}{0.31}*\text{VOLSTAF} + \underset{(0.256)}{0.513}*\text{NURSE}$$

$$* \underset{(0.275)}{0.483}*\text{SW}$$

$R^2 = 0.394$ $F(8,730)=59.2$ SER/LHS MEAN=2.48/2.1=1.18

5. $$\text{VOLSTAF} = \underset{(0.491)}{0.186} + \underset{(0.779)}{3.19}*\text{FFP} - \underset{(0.81)}{4.84}*\text{FFP}*\text{LEGSTAT} + \underset{(0.0122)}{0.08}*\text{FFP}*\text{LEGSTAT}*$$

$$\text{ENR} - \underset{(0.444)}{1.08}*(1 - \text{LEGSTAT})*\text{SUMDUM} + (\underset{(0.0361)}{0.143} - \underset{(0.05)}{0.09}*\text{FFP}$$

$$-\underset{(0.0255)}{0.129}*\text{LEGSTAT})*\text{DGSZ}*\text{AVSIZE}$$

$R^2 = 0.159$ $F(7,731)=19.7$ SER/LHS MEAN=4.03/1.78=2.26

6. $$\text{FTEPDSTF} = \underset{(0.139)}{0.911} + \underset{(0.0239)}{0.305}*\text{NCARE} + \underset{(0.0244)}{0.835}*\text{CFTE} - \underset{(0.226)}{0.767}*\text{FFP}$$

$$+ \underset{(0.0854)}{0.255}*\text{FFP}*\text{REQGRP}$$

$R^2 = 0.809$ $F(4,734)=776.$ SER/LHS MEAN=1.84/5.91=0.311

7. $$\text{EDMEAN} = \underset{(0.599)}{13.3} + \underset{(0.098)}{0.515}*\text{ZIPDUM4} - \underset{(0.0466)}{0.264}*\text{HRSOP} + \underset{(0.0000327)}{0.000285}*\text{MEDY} + \underset{(0.000124)}{0.000249}*\text{KH4}$$

$R^2 = 0.200$ $F(4,734)=45.8$ SER/LHS MEAN=1.2/13.6=0.0882

8. $$\text{EXMEAN} = \underset{(1.59)}{9.3} - \underset{(0.0331)}{0.0588}*\text{CFTE} + \underset{(0.018)}{0.202}*\text{YRSOPEN} + \underset{(0.257)}{0.597}*\text{ZIPDUM4} \underset{(0.246)}{-0\ 431}*\text{FFP} - \underset{(0.128)}{0.432}*\text{HRSOP} + \underset{(0.251)}{0.852}*\text{LEGSTAT} - \underset{(0.000085)}{0.000223}*\text{MEDY}$$

$R^2 = 0.209$ $F(7,731)=27.6$ SER/LHS MEAN=3.11/4.36=0.713

9. $$\text{KFTE} = \underset{(3.57)}{-51.2} + \underset{(0.337)}{4.91}*\text{HRSOP} + \underset{(0.696)}{2.64}*\text{SUMDUM} + (\underset{(0.0105)}{0.826} - \underset{(0.0146)}{0.042}*\text{FFP})*\text{FRNO} + \underset{(0.0267)}{0.977}*\text{GOVKIDS}$$

$R^2 = 0.918$ $F(5,733)=1640.$ SER/LHS MEAN=8.53/41.3=0.207

10. $$\text{NMEALS} = \underset{(11.8)}{33.0} - \underset{(6.13)}{15.9}*\text{CLS4} - \underset{(0.00122)}{0.00237}*\text{MEDY} + \underset{(0.0492)}{1.14}*\text{CAP}$$

$$+ \underset{(0.171)}{0.611}*\text{POORB}*\text{FAMB}/100 + \underset{(0.00523)}{0.0246}*\text{WELFIX2}*\text{GOVKIDS}$$

$$- \underset{(7.19)}{17.4}*\text{CLS6} + \underset{(4.62)}{14.4}*\text{USDAA}$$

$R^2 = 0.525$ F(7,731)=115.0 SER/LHS MEAN=46.6/81.5=0.572

11. $$\text{AVESAL1} = \underset{(46.2)}{58.7} + \underset{(11.9)}{51.0}*\text{CFTE}/\text{FTEPDSTF} - \underset{(12.2)}{28.3}*\text{FFP}*\text{LEGSTAT}$$

$$+ \underset{(0.0031)}{0.0179}*\text{MEDY} + \underset{(3.2)}{13.4}*\text{EDMEAN} + \underset{(19.3)}{61.7}*\text{REG2} + \underset{(1.14)}{2.04}*\text{EXMEAN}$$

$$+ \underset{(0.659)}{1.41}*\text{DGSZ}*\text{AVSIZE} + \underset{(15.2)}{80.4}*\text{GOVKIDS}/\text{ENR}$$

$R^2 = 0.186$ F(8,730)=20.9 SER/LHS MEAN=106./496.=0.214

12. $$\text{MONRENT} = \underset{(171.)}{-419.} + \underset{(0.971)}{7.39}*\text{CAP} + \underset{(0.0182)}{0.0576}*\text{MEDY} + \underset{(21.8)}{81.3}*\text{REQGRP}$$

$$+ \underset{(.00624)}{0.0196}*\text{FTCOST} - \underset{(55.5)}{309.}*\text{DONSPACE} - \underset{(56.7)}{276.}*\text{CLS2}$$

$$\underset{(58.2)}{-156.}*\text{CLS4} + \underset{(0.0123)}{0.0453}*\text{RNT100A} - \underset{(3.27)}{11.0}*\text{YRSOPEN}$$

$$\underset{(68.9)}{-170.}*\text{REG6} - \underset{(73.6)}{185.}*\text{REG9}$$

$R^2 = 0.39$ F(11,727)=42.3 SER/LHS MEAN=571./499.=1.14

13. MONSUP = -656. + 4.31*CAP + 2.96*NMEALS + 104.*SUMDUM
(186.) (1.06) (0.637) (57.9)

+ 0.0523*MEDY + 79.8*CFTE
(0.0191) (11.3)

R^2 = 0.405 F(5,733)=100.0 SER/LHS MEAN=727./760.=0.957

14. OTHCOST = -113. + 3.59*KFTE + 270.*PART - 96.6*SUMDUM + 168*LEGSTAT
(60.3) (1.06) (78.4) (57.3) (58.3)

+ 32.0*NCARE + 9.47*TRANP + 0.461*FFP*FRSHRS
(9.6) (2.58) (0.169)

R^2 = 0.147 F(7,731)=18.0 SER/LHS MEAN=717./280.=2.56

15. TRANP/ENR = -0.149 + 0.00543*GOVKIDS + 0.0343*LEGSTAT + 0.064*REG1
(0.0787) (0.000551) (0.0143) (0.0281)

+ 0.067*REG3 + 0.099*REG4 + 0.0523*REG5
(0.0258) (0.0254) (0.0209)

+ 0.0989*REG6 + 0.0133*MEDY/1000
(0.0273) (.00761)

R^2 = 0.134 F(8,730)=14.2 SER/LHS MEAN=0.185/0.09=2.31

16. $$\mathrm{UNDER3} = \underset{(128.)}{-718.} + \underset{(0.354)}{4.61}*\mathrm{FRNO} + \underset{(0.9)}{1.94}*\mathrm{GOVKIDS} + \underset{(12.1)}{70.7}*\mathrm{HRSOP}$$

$$+ \underset{(25.1)}{55.5}*\mathrm{SUMDUM}$$

$R^2 = 0.246$ F(4,734)=59.7 SER/LHS MEAN = 309./257.=1.20

17. $$\mathrm{KH3} = \underset{(95.7)}{-338.} + \underset{(0.261)}{6.95}*\mathrm{FRNO} + \underset{(0.673)}{11.6}*\mathrm{GOVKIDS} + \underset{(8.96)}{36.9}*\mathrm{HRSOP}$$

$R^2 = 0.552$ F(3,735)=302. SER/LHS MEAN=231./416.=0.555

18. $$\mathrm{KH5} = \underset{(93.1)}{-223.} + (\underset{(0.272)}{7.37} + \underset{(0.382)}{0.986}*\mathrm{FFP})*\mathrm{FRNO} + \underset{(0.699)}{9.17}*\mathrm{GOVKIDS}$$

$$+ \underset{(8.7)}{18.3}*\mathrm{HRSOP}$$

$R^2 = 0.585$ F(4,734)=259. SER/LHS MEAN=224./340.=0.659

19. $$\text{OVERS} = \underset{(110.)}{-675.} + \underset{(0.303)}{5.56}*\text{FRNO} + \underset{(0.77)}{5.14}*\text{GOVKIDS} + \underset{(10.4)}{52.4}*\text{HRSOP} + \underset{(21.5)}{117.}*\text{SUMDUM}$$

R^2 = 0.367 F(4,734)=106. SER/LHS MEAN=264.0/181.=1.46

20. $$\text{AVSIZE} = \text{DGSZ}*[\underset{(1.08)}{8.65} + \underset{(0.0205)}{0.087}*(\text{KH3} + \text{KH4})/40 + \underset{(0.0173)}{0.0393}*\text{FFP}*(\text{KH3} + \text{KH4})/40$$

$$+ \underset{(0.0293)}{0.0941}*\text{LEGSTAT}*(\text{KH3} + \text{KH4})/40 + \underset{(0.0418)}{0.134}*\text{ENR}/\text{REQGRP}$$

$$- \underset{(0.767)}{1.55}*\text{LEGSTAT}] + \underset{(0.03)}{1.02}*(1-\text{DGSZ})*(\text{KH3} + \text{KH4})/40$$

R^2 = 0.535 F(6,732)=140. SER/LHS MEAN=5.84/14.5=0.403

21. $$\text{HRSOP} = \underset{(0.183)}{10.7} - \underset{(0.086)}{0.329}*\text{CLS2} + \underset{(0.136)}{0.226}*\text{CLS3} - \underset{(0.0961)}{0.257}*\text{CLS4}$$

$$+ \underset{(0.158)}{0.285}*\text{CLS5} - \underset{(0.00481)}{0.00925}*\text{WELFIX2} + \underset{(0.00112)}{0.00293}*\text{ENR}$$

$$+ \underset{(0.0126)}{0.0369}*\text{ACSR}$$

R^2 = 0.0896 F(7,731)=10.3 SER/LHS MEAN=0.922/10.7=0.0862

22. $$\mathrm{SRSHRS} = \underset{(2.95)}{20.6} + \underset{(0.0712)}{1.25}*\mathrm{SR3}*\mathrm{KH3} + \underset{(0.0738)}{0.968}*\mathrm{SR4}*\mathrm{KH4} + \underset{(0.0832)}{0.937}*\mathrm{SR5}*\mathrm{KH5}$$

$$+ \underset{(0.00551)}{0.083}*\mathrm{OVER5}$$

$R^2 = 0.792$ F(4,734)=700. SER/LHS MEAN=48.0/143.=0.336

23. $$\mathrm{RICHB} = \mathrm{FFP}*(\underset{(10.1)}{-8.99} + \underset{(0.06)}{0.285}*(100-\mathrm{POORB}) + \underset{(0.00101)}{0.00141}*\mathrm{MEDY}$$

$$- \underset{(0.0554)}{0.109}*100*\mathrm{GOVKIDS}/\mathrm{ENR}) + (1-\mathrm{FFP})*(\underset{(7.09)}{-8.9} + \underset{(2.14)}{9.31}*\mathrm{LEGSTAT}$$

$$+ \underset{(0.00107)}{0.00194}*\mathrm{MEDY} + \underset{(0.6)}{1.5}*\mathrm{COLPCT})$$

$R^2 = 0.3$ F(7,731)=31.2 SER/LHS MEAN=22.3/26.7=0.835

24. $$\mathrm{POORB} = \mathrm{FFP}*[100*\mathrm{GOVKIDS}/\mathrm{ENR} + \underset{(8.41)}{3.48} - \underset{(0.000854)}{0.00293}*\mathrm{MEDY}$$

$$+ (\underset{(0.0437)}{0.503} - \underset{(0.0957)}{0.252}*(\mathrm{LEGSTAT})*100*\mathrm{FRNO}/\mathrm{ENR} + \underset{(7.26)}{11.6}*\mathrm{LEGSTAT}]$$

$$+ (1-\mathrm{FFP})*[\underset{(4.09)}{-10.3} + \underset{(5.18)}{8.88}*\mathrm{LEGSTAT} - \underset{(0.071)}{0.195}*\mathrm{LEGSTAT}*(100-\mathrm{RICHB})$$

$$+ \underset{(0.0542)}{0.383}*(100-\mathrm{RICHB})]$$

$R^2 = 0.51$ F(8,730)=95.1 SER/LHS MEAN=19.1/24.8=0.770

25. $\mathrm{KH4} = \mathrm{KFTE}*40 - \mathrm{UNDER3} - \mathrm{KH3} - \mathrm{KH5} - \mathrm{OVER5}$

26. REQGRP = UNDER3/600 + KH3/600 + KH4/800 + KH5/800
+ OVER5/1000

27. FRSHRS = UNDER3/4 + KH3/5 + KH4/7 + KH5/7 + OVER5/20

28. SCSR = 40*KFTE/SRSHRS

29. FCSR = 40*KFTE/FRSHRS

30. ACSR = 0.88*KFTE/CFTE

31. SALEXP = AVESAL1*FTEPDSTF

```
32.  MONCOST1 = SALEXP + MONRENT + MONSUP + OTHCOST


33.  COSTPK = MONCOST1/KFTE


34.  ENR = FRNO + GOVKIDS


35.  SOK = IF AMSR GT - 0.125 THEN 1 ELSE 0


36.  FOK = IF AMFR GT - 0.125 THEN 1 ESLE 0


37.  AMFR = CFTE - 0.88*FRSHRS/40


38.  AMSR = IF STATE EQ 34 THEN CFTE - SRSHRS/40
             ELSE CFTE - 0.88*SRSHRS/40
```

39. $$\text{PEX} = \underset{(0.026)}{0.0667} + \underset{(0.0412)}{0.172}*\text{FFP} - \underset{(0.0531)}{0.14}*\text{FFP}*\text{LEGSTAT} + \underset{(0.041)}{0.128}*\text{NURSE}$$

$$+ \underset{(0.22)}{0.403}*\text{NCARE/ENR} + \underset{(0.00416)}{0.00751}*\text{DCVISIT} + \underset{(0.00066)}{0.0021}*\text{POORB}$$

$R^2 = 0.112$ $\quad$ F(6,732)=15.4 $\quad$ SER/LHS MEAN=0.401/0.234=1.71

40. $$\text{DEX} = \underset{(0.0993)}{-0.294} + \underset{(0.0395)}{0.223}*\text{FFP} - \underset{(0.0505)}{0.154}*\text{FFP}*\text{LEGSTAT} + \underset{(0.0395)}{0.116}*\text{NURSE}$$

$$+ \underset{(0.00404)}{0.0125}*\text{DCVISIT} + \underset{(0.000634)}{0.00212}*\text{POORB} + \underset{(0.0102)}{0.0363}*\text{MEDY}/1000$$

$R^2 = 0.153$ $\quad$ F(6,732)=22.0 $\quad$ SER/LHS MEAN=0.388/0.227=1.71

41. $$\text{PARSERV3} = \underset{(0.0402)}{0.249} + \underset{(0.0406)}{0.272}*\text{FFP} - \underset{(0.0355)}{0.0994}*\text{LEGSTAT} + \underset{(0.000733)}{0.00204}$$

$$*\text{POORB} + \underset{(0.00461)}{0.00874}*\text{CFTE}$$

$R^2 = 0.165$ $\quad$ F(4,734)=36.2 $\quad$ SER/LHS MEAN=0.450/0.406=1.11

42. $$\mathrm{SW} = \underset{(0.0255)}{0.00908} + \underset{(0.0332)}{0.06}*\mathrm{FFP}*(1-\mathrm{LEGSTAT}) + \underset{(0.000534)}{0.00149}*\mathrm{POORB}$$

$$+ \underset{(0.0263)}{0.0626}*\mathrm{ZIPDUM4} + \underset{(0.027)}{0.0638}*\mathrm{PARSERV3} + \underset{(0.0481)}{0.111}*\mathrm{REG1}$$

$R^2 = 0.065$ $F(5,733)=10.2$ SER/LHS MEAN=0.335/0.138=2.43

43. $$\mathrm{NURSE} = \underset{(0.0209)}{0.149} + \underset{(0.018)}{0.0555}*(\mathrm{PEX} + \mathrm{DEX}) - \underset{(0.027)}{0.0451}*\mathrm{LEGSTAT}$$

$R^2 = 0.02$ $F(2,736)=7.47$ SER/LHS MEAN=0.358/0.154=2.32

44. $$\mathrm{PART} = \underset{(0.0296)}{0.191} + \underset{(0.0438)}{0.185}*\mathrm{FFP} - \underset{(0.0313)}{0.124}*\mathrm{LEGSTAT} - \underset{(0.05445)}{0.176}*\mathrm{FFP}*\mathrm{LEGSTAT}$$

$$+ \underset{(0.00141)}{0.00365}*\mathrm{GOVKIDS} - \underset{(0.0424)}{0.129}*\mathrm{DK40} - \underset{(0.0261)}{0.06}*\mathrm{ZIPDUM4}$$

$$+ \underset{(0.00657)}{0.0177}*\mathrm{FFP}*\mathrm{NCARE}$$

$R^2 = 0.201$ $F(7,731) = 26.3$ SER/LHS MEAN=0.33/0.161=2.05

45. $$\text{REVIEW} = \underset{(0.0425)}{0.344} + \underset{(0.0511)}{0.111}{*}\text{FFP} - \underset{(0.0391)}{0.13}{*}\text{LEGSTAT} - \underset{(0.0672)}{0.13}{*}\text{FFP}{*}\text{LEGSTAT}$$

$$+ \underset{(0.00154)}{0.00542}{*}\text{GOVKIDS} - \underset{(0.000468)}{0.001}{*}\text{FRNO} - \underset{(0.0323)}{0.0717}{*}\text{ZIPDUM4}$$

$$+ \underset{(0.00366)}{0.0195}{*}\text{VOLSTAF}$$

$R^2 = 0.199$ $F(7,731)=25.9$ SER/LHS MEAN=0.409/0.292=1.4

46. $$\text{USDAA} = (1 - \text{LEGSTAT}){*}[\underset{(0.0466)}{0.421} + \underset{(0.0473)}{0.206}{*}\text{FFP} + \underset{(0.00142)}{0.00879}{*}\text{GOVKIDS}$$

$$+ \underset{(0.0205)}{0.103}{*}(1 - \text{FFP}){*}\text{POORB}{*}\text{ENR}/1000 - \underset{(0.0365)}{0.0832}{*}\text{ZIPDUM4}$$

$$- \underset{(0.0644)}{0.121}{*}(1 - \text{FFP}){*}\text{DNEW} - \underset{(0.000565)}{0.00141}{*}\text{ENR}]$$

$R^2 = 0.433$ $F(6,732)=93.2$ SER/LHS MEANS=0.34/0.28=1.21